UCLA
UCLA Electronic Theses and Dissertations

Title
Studying the Evolution of Emerging Viruses using Genome Wide Analysis

Permalink
https://escholarship.org/uc/item/8fv9v596

Author
Wang, Lulan

Publication Date
2019

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Studying the Evolution of Emerging Viruses

using Genome Wide Analysis

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Molecular Biology

by

Lulan Wang

2019

ABSTRACT OF THE DISSERTATION

Studying the Evolution of Emerging Viruses
using Genome Wide Analysis

by

Lulan Wang

Doctor of Philosophy in Molecular Biology

University of California, Los Angeles, 2019

Professor Genhong Cheng, Chair

In arms-race between the human host and viral pathogens, the favor often belongs to the pathogens. I have investigated emerging viruses by using meta-genomic data and system-wide mutagenesis of Zika and Influenza virus.

Initially isolated from the Zika forest at Uganda in 1947. Zika virus (ZIKV) has been spreading in Africa, Asia and Asian-Pacific islands. It is not until late 2015 that Zika virus has emerged as a virus with significant public health impact, with rapidly accumulating evidence of a causal relationship with fetal neurologic birth defects. First reported from Northeast of Brazil in May 2015, followed by Colombia and Suriname in October 2015. Over the next couple of years, the virus has spread over 130 countries and infected tens of millions of people globally. It has been hypothesized that the virus may have recently evolved to become more neurotropic, to exhibit increased replicative capacity, and/or to become more transmissible to humans, but causal support for these possibilities is outstanding. To gain a better understanding of the molecular evolution of the virus, we have developed a novel genetic footprint method, as well as targeted structural modeling, on all known full-length ZIKV available to date. We were able to classify the ZIKV strains into ten groups and trace the spread and evolution of ZIKV in different parts of Brazil and other countries. Our analysis further pointed to the possibility that different groups of ZIKV may trigger different levels of host immune responses and cause distinct levels of pathogenesis.

Vaccination is one of the greatest accomplishments in modern medicine and has effectively eradicated several life-threatening viruses in the past. A vaccine is essentially the injection of an attenuated pathogen, allowing the immune system to recognize and generate protective antibodies, which could serve future protection when re-exposed to the same pathogen. Over the years, the development of vaccine has followed an empirical paradigm, which has been successful for a few viruses, but not others. We began to recognize that viruses have developed their own complex strategies to combat vaccine-induced immune recognition. Viruses like influenza and HIV have developed powerful mutational strategies, that requires rational vaccine-design strategies to reduce or eradicate.

Influenza A virus is a major public health problem, infecting as many as 500 million people a year worldwide. Unfortunately, the current vaccines are highly inefficient in its cross-protection. It is urgent that we uncover new strategies for generating a universal vaccine. Here we combined high-coverage and unbiased transposon mutagenesis of influenza virus with a rapid high-throughput screening for attenuation to generate W7-791, a live attenuated mutant virus strain. W7-791 produced only a transient asymptomatic infection in adult and neonatal mice even at doses 100-fold higher than the LD_{50} of the parent strain. A single administration of W7-791 conferred full protection to mice against lethal challenge with H1N1, H3N2, and H5N1 strains, and improved viral clearance in ferrets. Furthermore, W7-791-immunized mice conferred heterologous protection, indicating a role for T cell-mediated immunity.

The dissertation of Lulan Wang is approved.

Peter John Bradley

Robert L Modlin

Hong Zhou

Genhong Cheng, Committee Chair

University of California, Los Angeles

2019

DEDICATION

To my parents, Huizhu Chen and Shiping Wang,

For their continued support to my scientific career.

To my mentor, Dr. Genhong Cheng,

For his guidance with vision and optimism.

To my wife and our son, Zhenzi Pu, Harris Wang

For her love, encouragement and support

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 3

LIST OF ACRONYMS

Ab   Antibody
C   capsid (Zika)
DENV   Dengue virus
DNA   Ribonucleic acid
DOHMH   NYC Department of Health and Mental Hygiene
E   Envelope (Zika)
ER   Endoplasmic reticulum
FSM   Zika Micronesia/2007 strain
GBS   Guillain-Barré syndrome
H/PF/2013   Zika French Polynesia/2013 strain
HA   Hemagglutinin (influenza)
HAI assay   Hemagglutinin inhibition assay
IbH-30656   Zika Nigeria/1968 strain
LAIV   Live attenuated influenza vaccine
LAV   Live attenuated vaccine
M gene   Matrix gene (influenza)
MDCK   Madin-Darby canine kidney
MEGA   Molecular Evolutionary Genetic Analysis
ML   Maximum Likelihood
NA   Neuraminidase (influenza)
NCBI   The National Center for Biotechnology Information
NGS   Next generation sequencing
NIAID   National Institute of Allergy and Infectious Diseases
NS1   Non-structural protein 1 (Zika)
ORF   Open reading frame
PDB   Protein database
prM   Pre-membrane precursor
RLU   Relative light units
RNA   Deoxyribonucleic acid
SNPs   Single Nucleotide Polymorphism
TBC   The template-binding and catalytic channel
TCR   T cell receptor
UTR   Untranslated regions
ViPR   Pathogen Database and Analysis Resource
WHO   World Health Organization
WSN   Influenza A/WSN/1933 strain
WT   wild-type
ZIKV   Zika Virus

ACKNOWLEDGMENTS

First and foremost, I would like to express my most sincere gratitude to my supervisor, Professor Genhong Cheng, for believing and supporting my pursuit of research throughout the years. He has provided an excellent laboratory atmosphere, with scientific freedom to explore various questions and interest. His enthusiasm and curiosity for science are key motivators throughout my PhD. Whenever I needed help, he has always been willing to go for the extra mile. I am truly thankful for his selfless dedication to both my personal and academic development. I cannot think of a better supervisor to have. Over the years, I have learned vital skills of scientific thinking, grant writing and scientific communications. I owe a great debt of gratitude to Genhong.

I would like to thank the members of my thesis committee, Drs. Peter J. Bradley, Robert R. Modlin, Otto O. Yang and Z. Hong Zhou who's for proving me with valuable feedback for my thesis and suggestions to scientific research.

I am grateful for the support; guidance of the many friends and I have made over the years in the Cheng Lab. I am appreciative of the diversity of interests both scientific. In particular, I would like to thank Drs Suyang Liu, Yao Wang, Saba Aliyari, Feng Ma, Natalie Quanquin, Maxim Chapon and Kislay Parvatiyar for their contributions to my graduate work as well as their time and advice during our many productive conversations over the years. I would also like to thank my research collaborators, Drs. Aiping Wu, Chunfeng Li, Taijiao Jiang, Ouyang Song, Yingfang Liu, Jane Deng, Xiao Li and Karin Nielsen. I would like to also acknowledge four undergraduate students, Stephan Lam, William Lee, Cat Van Tran and Angela Zou who have provided a great opportunity for me to train new scientists and without whom much of this work could not be done.

I would like to thank my previous research mentor, professor Paul Zhou who have been instrumental in my development as a scientist and have profoundly shaped my scientific interests.

A special thanks to friends and family outside the laboratory, Drs Xianzhi Lin, Yulin Zeng and Wanlu Liu.

UNIVERSITY OF CALIFORNIA, LOS ANGELES
Institute of Molecular Biology
Research Assistant/Graduate Student
Adviser: Dr. Genhong Cheng

INSTITUT PASTEUR, CHINESE ACADEMY OF SCIENCES
Unit of Anti-Viral Immunity and Genetic Therapy
Research Assistant
Adviser: Dr. Paul Zhou

FACULTY OF SCIENCES, MAHIDOL UNIVERSITY
Department of Pathology
Bachelor Student Graduate Student
Adviser: Dr. Wannee Jiraungkoorskul

PUBLICATIONS

1. Functional Genomics Reveals Linkers Critical for Influenza Virus Polymerase.
Wang L, Wu A, Wang YE, Quanquin N, Li C, Wang J, Chen HW, Liu S, Liu P, Zhang H, Qin FX, Jiang T, Cheng G. J Virol. 2015 Dec 30;90(6):2938-47.

2. Generation of a Live Attenuated Influenza Vaccine that Elicits Broad Protection in Mice and Ferrets.
Wang L, Liu SY, Chen HW, Xu J, Chapon M, Zhang T, Zhou F, Wang YE, Quanquin N, Wang G, Tian X, He Z, Liu L, Yu W, Sanchez DJ, Liang Y, Jiang T, Modlin R, Bloom BR, Li Q, Deng JC, Zhou P, Qin FX, Cheng G. Cell Host Microbe. 2017 Mar 8;21(3):334-343.

3. From Mosquitos to Humans: Genetic Evolution of Zika Virus.
Wang L, Valderramos SG, Wu A, Ouyang S, Li C, Brasil P, Bonaldo M, Coates T, Nielsen-Saines K, Jiang T, Aliyari R, Cheng G. Cell Host Microbe. 2016 May 11;19(5):561-5.

4. Tracking the Evolution of and Spread of Zika Virus by Footprint Analysis.
Wu A, **Wang L**, Chapon M, Liu D, Li C, Ren W, Fu J, Deocharan B, Quanquin N, Gaw S, Aliyari S, Luo J, Zhu W, Tran CV, Jung J, Nielsen-Saines K, Jiang T, Qin CF, Rakeman J, Cheng G. Under Review.

5. Potential for treatment and a Zika virus vaccine.
Quanquin N, **Wang L**, Cheng G. Curr Opin Pediatr. 2017 Feb;29(1):114-121. Review

6. Integrating computational modeling and functional assays to decipher the structure-function relationship of influenza virus PB1 protein. Li C, Wu A, Peng Y, Wang J, Guo Y, Chen Z, Zhang H, Wang Y, Dong J, **Wang L**, Qin FX, Cheng G, Deng T, Jiang T. Sci Rep. 2014 Nov 26;4:7192.

7. Screening for Novel Small-Molecule Inhibitors Targeting the Assembly of Influenza Virus Polymerase Complex by a Bimolecular Luminescence Complementation-Based Reporter System.
Li C, Wang Z, Cao Y, **Wang L**, Ji J, Chen Z, Deng T, Jiang T, Cheng G, Qin FX. J Virol. 2017 Feb 14;91(5). pii: e02282-16.

8. Cryo-EM structure of influenza virus RNA polymerase complex at 4.3 Å resolution.
Chang S, Sun D, Liang H, Wang J, Li J, Guo L, Wang X, Guan C, Boruah BM, Yuan L, Feng F, Yang M, **Wang L**, Wang Y, Wojdyla J, Li L, Wang J, Wang M, Cheng G, Wang HW, Liu Y. Mol Cell. 2015 Mar 5;57(5):925-935.

9. Systematic characterization of ISGs that can inhibit influenza virus replication
Li C, Wang Y, Wang Y, **Wang L**, Qin X, Cheng G. Ninth Annual National Immunology Meeting Publication. 2014

CHAPTER 1

INTRODUCTION

1. BACKGROUND

Emerging virus infection is one of the largest global socio-economic burden. It has always taken the world by surprise. To date, there has been no predictive intervention that can be used to prevent the next epidemic outbreak. The emerging of Zika and recombined Influenza virus in the recent years has combined over millions of human infections worldwide. Therefore, understanding viral evolution and mutation is critical for the community as well as for the development of therapeutics and vaccines.

1.1 ZIKA VIRUS

Zika virus (ZIKV), was first discovered and isolated in Zika forest, Uganda in 1947. It has only caused sporadic disease throughout Africa and Asia until the 2007 Micronesia outbreak, followed by 2013 French Polynesia outbreaks (1). Faria et al, shown that the virus has travelled to Brazil during late 2013 or early 2014 (2). The rapid expansion of its geographic range and the increase in severe pathogenicity first noted in the 2015-2016 Brazilian outbreak have raised questions regarding the molecular evolution of this virus and the demographic population of the Americas. Previously believed to cause only mild disease, mounting evidence points to the capacity of ZIKV to cause neuropathology, including abnormal fetal brain development and Guillain-Barré syndrome (GBS) (2). In addition to the rise of associated disorders, novel modes of ZIKV transmission have been reported, including maternal-fetal transmission (3) and sexual transmission (4). In February 2016, the World Health Organization (WHO) declared a Public Health Emergency of International Concern due to observations that infection with ZIKV during pregnancy led to congenital anomalies, especially microcephaly.

The mechanisms underlying the ZIKV evolution from a benign virus to highly pathogenic strains, and immune system responses against ZIKV infection, have yet to be elucidated. ZIKV is

2

a flavivirus closely related to dengue virus (DENV), encoded by an approximately 10.8kb long single-stranded positive-sense RNA molecule. The 5' and 3' untranslated (UTR) regions flank a large open reading frame whose resulting single polyprotein is cleaved into the capsid (C), pre-membrane precursor (prM), and envelope (E) structural proteins, and the non-structural proteins NS1, NS2A, NS2B, NS3, NS4A, 2K, NS4B, and NS5 (5). To gain a better understanding of the molecular evolution of the virus, we have performed a detailed phylogenetic and genetic analyses, as well as targeted structural modeling, on the known full-length genome nucleotide sequences from 41 strains (30 human isolates, ten mosquito isolates, and one monkey isolate) in early 2016 (6). We found that all ZIKV strains isolated from the South American outbreaks shared a greater genome sequence homology with the ZIKV mosquito isolate from Malaysia in 1966, than with human or mosquito isolates from Africa. Prior genetic and phylogenetic analyses classified ZIKV strains into either the African or Asian lineage, and over the last few decades, all epidemics have been related to the Asian lineage (7-9). Despite circulating throughout both Africa and Asia for the latter half of the 20th century, ZIKV infections were not reported to cause significant clinical outcomes until recently. Our studies raise the interesting question of how this virus, which was until recently considered relatively benign, has emerged into highly pathogenic and dangerous strains. Our phylogenetic analysis has revealed numerous sequence variations in the ZIKV genomes between the African and Asian lineages, as well as among different strains within the Asian lineage. Our modeling studies suggest that these sequence variations could mediate specific changes in the prM protein, which led to the recent discovery that the S139N single amino acid change significantly increased the virulence of the ZIKV strains (10). In Chapter 2.1, we will first demonstrate our early findings during Zika outbreak. Following the initial research, we have taken advantage of a genetic footprint method to further track the evolution of ZIKV strains. In Chapter 2.2, we have identified ZIKV viral clades which could help with epidemiological tracking of outbreaks. In addition to genotyping, we demonstrated the importance of recessive SNPs in

NGS data, which should be utilized to aid genome tracing of ZIKV. And finally, in Chapter 2.3, we have demonstrated therapeutic and vaccine development against ZIKV.

1.2 INFLUENZA VIRUS

Influenza A virus is a major public health problem, infecting as many as 500 million people a year worldwide and leading to more than 500,000 deaths (11). During a pandemic, the mortality rate is significantly higher, as demonstrated by the 40-50 million worldwide deaths caused by the 1918 "Spanish Flu". In the United States, 5-10% of the population is infected by influenza virus in a typical year, resulting in ~220,000 hospitalizations and ~36,000 deaths (12-13).

Influenza virus has a segmented genome containing eight genes encoding 11 proteins. It utilizes an RNA-dependent RNA polymerase (RdRp) which consists of the PA (polymerase acidic), PB1 (polymerase basic 1) and PB2 (polymerase basic 2) subunits. It has been the focus of many researchers as they were more conserved therapeutic target than external glycoproteins. However, its multiple functions and how they relate to specific structures within the heterotrimeric complex have not yet been fully elucidated. To study the structure-function of the polymerase complex, we have utilized a transposon-based insertional mutagenesis, As shown in Chapter 3.1, we generated high-density mutant libraries for PA, PB1 and PB2 influenza gene segments, with coverage at almost every nucleotide position. The insertions include a particular 15 nt sequence that does not cause truncations or frame-shifts, and can be used to effortlessly locate the mutation sites of each clone. Entire libraries of these mutant genes can be rapidly characterized all at once with high-throughput genotyping and next generation sequencing. This allows us to simultaneously measure the rate of growth attenuation of thousands of clones over multiple passages in vitro. Our study has provided new insights to the structure and functional relationship of the influenza polymerase complex.

In addition to understanding the biology of influenza virus, our ultimate objectives of studying influenza are to generate novel and effective live attenuated influenza vaccines with broad protection against multiple strains. The most widely used influenza vaccine, consisting of 3-4 strains of killed influenza virus, elicits Ab responses only to the viral coat proteins hemagglutinin (HA) and neuraminidase (NA), which frequently mutate to create new subtypes of virus, rendering the previous year's vaccine ineffective. A critical measure for an effective influenza vaccine is to determine whether it generates sufficient broadly neutralizing antibody in an individual. Therefore, we studied the composition of individuals Ab population, and learned that it differs based on the robustness of their immune system and their previous exposures to vaccines and infections. Even knowing their clinical history, it may be difficult to predict whether a person is actually protected from a particular influenza strain. This is further complicated by the fact that current influenza vaccines do not effectively cross-protect against different strains, leaving us vulnerable to emerging new influenza viruses such as H5N1, H7N9 and most recently, H5N8. A live attenuated influenza vaccine (LAIV) is the most promising model for inducing broad protection, for in addition to eliciting humoral immunity to surface antigens, it can trigger cell-mediated responses against the more conserved targets within the viral particle (14). Unfortunately, current methods of generating LAIVs are highly inefficient, and have so far not generated a vaccine that can effectively cross-protect against multiple strains (15-16). Given how vulnerable we are to infection by influenza subtypes that were not predicted to circulate, or that have mutated and against which we may have no protection, it is urgent that we uncover new strategies for generating a universal vaccine. In Chapter 3.2, we have developed a novel rapid influenza attenuation screening system, and identified an insertion mutant that is both safe and can effectively protect mice and ferrets from lethal challenge against multiple group 1 and group 2 influenza A viruses.

REFERENCE

1. Broutet N, Krauer F, Riesen M, Khalakdina A, Almiron M, Aldighieri S, Espinal M, Low N, Dye C. 2016. Zika Virus as a Cause of Neurologic Disorders. The New England journal of medicine 374:1506-1509.

2. Brasil P, Pereira JP, Jr., Moreira ME, Ribeiro Nogueira RM, Damasceno L, Wakimoto M, Rabello RS, Valderramos SG, Halai UA, Salles TS, Zin AA, Horovitz D, Daltro P, Boechat M, Raja Gabaglia C, Carvalho de Sequeira P, Pilotto JH, Medialdea-Carrera R, Cotrim da Cunha D, Abreu de Carvalho LM, Pone M, Machado Siqueira A, Calvet GA, Rodrigues Baiao AE, Neves ES, Nassar de Carvalho PR, Hasue RH, Marschik PB, Einspieler C, Janzen C, Cherry JD, Bispo de Filippis AM, Nielsen-Saines K. 2016. Zika Virus Infection in Pregnant Women in Rio de Janeiro. The New England journal of medicine 375:2321-2334.

3. Calvet G, Aguiar RS, Melo AS, Sampaio SA, de Filippis I, Fabri A, Araujo ES, de Sequeira PC, de Mendonca MC, de Oliveira L, Tschoeke DA, Schrago CG, Thompson FL, Brasil P, Dos Santos FB, Nogueira RM, Tanuri A, de Filippis AM. 2016. Detection and sequencing of Zika virus from amniotic fluid of fetuses with microcephaly in Brazil: a case study. The Lancet. Infectious diseases 16:653-660.

4. Hills SL, Russell K, Hennessey M, Williams C, Oster AM, Fischer M, Mead P. 2016. Transmission of Zika Virus Through Sexual Contact with Travelers to Areas of Ongoing Transmission - Continental United States, 2016. MMWR. Morbidity and mortality weekly report 65:215-216.

5.   Kuno G, Chang GJ. 2007. Full-length sequencing and genomic characterization of Bagaza, Kedougou, and Zika viruses. Archives of virology 152:687-696.

6.   Wang L, Valderramos SG, Wu A, Ouyang S, Li C, Brasil P, Bonaldo M, Coates T, Nielsen-Saines K, Jiang T, Aliyari R, Cheng G. 2016. From Mosquitos to Humans: Genetic Evolution of Zika Virus. Cell host & microbe 19:561-565.

7.   Haddow AD, Schuh AJ, Yasuda CY, Kasper MR, Heang V, Huy R, Guzman H, Tesh RB, Weaver SC. 2012. Genetic characterization of Zika virus strains: geographic expansion of the Asian lineage. PLoS Negl Trop Dis 6:e1477.

8.   Faye O, Freire CC, Iamarino A, Faye O, de Oliveira JV, Diallo M, Zanotto PM, Sall AA. 2014. Molecular evolution of Zika virus during its emergence in the 20(th) century. PLoS neglected tropical diseases 8:e2636.

9.   Lanciotti RS, Lambert AJ. 2016. Phylogenetic Analysis of Chikungunya Virus Strains Circulating in the Western Hemisphere. Am J Trop Med Hyg.

10.   Yuan L, Huang XY, Liu ZY, Zhang F, Zhu XL, Yu JY, Ji X, Xu YP, Li G, Li C, Wang HJ, Deng YQ, Wu M, Cheng ML, Ye Q, Xie DY, Li XF, Wang X, Shi W, Hu B, Shi PY, Xu Z, Qin CF. 2017. A single mutation in the prM protein of Zika virus contributes to fetal microcephaly. Science (New York, N.Y.) 358:933-936.

11.   WHO. Influenza: Fact sheets 2003.

12.   Patterson KD, Pyle GF. The geography and mortality of the 1918 influenza pandemic. Bull Hist Med. 1991;65(1):4-21. PubMed PMID: 2021692.

13.   Taubenberger JK, Reid AH, Janczewski TA, Fanning TG. Integrating historical, clinical and molecular genetic data in order to explain the origin and virulence of the 1918 Spanish influenza virus. Philos Trans R Soc Lond B Biol Sci. 2001;356(1416):1829-39. doi: 10.1098/rstb.2001.1020. PubMed PMID: 11779381; PMCID: PMC1088558.

14.   Pica N, Palese P. Toward a universal influenza virus vaccine: prospects and challenges. Annu Rev Med. 2013;64:189-202. doi: 10.1146/annurev-med-120611-145115. PubMed PMID: 23327522.

15.   He XS, Holmes TH, Zhang C, Mahmood K, Kemble GW, Lewis DB, Dekker CL, Greenberg HB, Arvin AM. Cellular immune responses in children and adults receiving inactivated or live attenuated influenza vaccines. J Virol. 2006;80(23):11756-66. doi: 10.1128/JVI.01460-06. PubMed PMID: 16971435; PMCID: 1642596.

16.   Hoft DF, Babusis E, Worku S, Spencer CT, Lottenbach K, Truscott SM, Abate G, Sakala IG, Edwards KM, Creech CB, Gerber MA, Bernstein DI, Newman F, Graham I, Anderson EL, Belshe RB. Live and inactivated influenza vaccines induce similar humoral responses, but only live vaccines induce diverse T-cell responses in young children. J Infect Dis. 2011;204(6):845-53. doi: 10.1093/infdis/jir436. PubMed PMID: 21846636; PMCID: 3156924.

17.   Guo H, Baker SF, Martinez-Sobrido L, Topham DJ. Induction of CD8 T cell heterologous protection by a single dose of single-cycle infectious influenza virus. J Virol. 2014;88(20):12006-16. doi: 10.1128/JVI.01847-14. PubMed PMID: 25100831; PMCID: 4178714.

CHAPTER 2

GENETIC EVOLUTION OF ZIKA VIRUS

2.1 FROM MOSQUITOS TO HUMANS: GENETIC EVOLUTION OF ZIKA VIRUS

ABSTRACT

Initially isolated in 1947, Zika virus (ZIKV) has recently emerged as a virus with significant public health impact. Sequence analysis of all 41 known ZIKV RNA open reading frames to date indicates that ZIKV has undergone significant changes in both protein and nucleotide sequences during the past half century.

2.1.a INTRODUCTION

Zika virus (ZIKV), discovered in 1947, had caused sporadic disease throughout Africa and Asia until the 2007 Micronesia and 2013 French Polynesia outbreaks (Broutet et al., 2016). The rapid expansion of geographic range and increase in severe pathogenicity first noted in 2015-2016 Brazilian outbreak has raised questions regarding the molecular evolution of this virus. Previously believed to cause only mild disease, mounting evidence points to the capacity of ZIKV to cause neuropathology, including disorders of fetal brain development and Guillain-Barre syndrome (Broutet et al., 2016). In addition to the rise of associated disorders, novel modes of ZIKV transmission have been reported, including maternal-fetal transmission (Brasil et al., 2016; Calvet et al., 2016; Sarno et al., 2016) and sexual transmission (Hills et al., 2016).

ZIKV is a flavivirus closely related to dengue virus (DENV). Its genome is a single-stranded positive sense RNA molecule of approximately 10800 base pairs, closely related to dengue virus (DENV). A single open reading frame (ORF) is flanked by 5' and 3' untranslated regions (UTRs). The resulting single polyprotein is cleaved into the structural proteins capsid (C), pre-membrane precursor (prM), and envelope (E), and the non-structural proteins NS1, NS2A, NS2B, NS3, NS4A, 2K, NS4B, and NS5 (Kuno and Chang, 2007). Prior genetic and phylogenetic analyses have identified two main ZIKV lineages: African and Asian, and the recent 21st century epidemics have been traced to the Asian lineage (Faye et al., 2014; Haddow et al., 2012; Lanciotti and Lambert, 2016). Despite circulating throughout Africa and Asia for the latter half of the 20th century, ZIKV infections were not associated with significant human pathology until now. The reasons for this are obscure. It has been hypothesized that the virus may have recently evolved to become more neurotropic, to exhibit increased replicative capacity, and/or to become more transmissible to humans, but causal support for these possibilities is outstanding. To gain a better understanding of the molecular evolution of the virus, we performed detailed phylogenetic and genetic analyses,

as well as targeted structural modeling, on all known full-length ORFs of ZIKV available to date (with an emphasis on the recent human strains).

2.1.b RESULTS

Nucleotide sequences from 41 strains were included in the analysis: 30 human isolates (including two newly-reported here), ten mosquito isolates, and one monkey isolate. All sequences greater than 10.2 kb in length were included in the analysis to consider all complete sequences of ORFs. The strains analyzed, including accession numbers, year, location, and source of isolation is listed in **Table S1**. We first investigated the phylogenetic relationships for all full length ORF sequences by using maximum likelihood (ML) mapping method with 1000 replicates. Consistent with prior reports, there were two major lineages of ZIKV, African and Asian **(Figure 1A**). Interestingly, the African lineage contained eight mosquito isolates, whereas P6-740 (Malaysia/1966) was the sole mosquito isolate in the Asian lineage. All the contemporary human strains share greater sequence homology to the P6-740 than IbH-30656 (Nigeria/1968), suggesting that the ZIKV strains in the recent human outbreak are evolved from the Asian lineage, which is anchored by P6-740 **(Figure S1)**. All of the human strains identified in the 2015-16 epidemic appear to be more closely related to the H/PF/2013 strain (French Polynesia/2013) than the FSM strain (Micronesia/2007), suggesting that perhaps these two variants have evolved in parallel from a common ancestor. Furthermore, compared to the mosquito strain, 435 and 446 nucleotide changes are evident in FSM and H/PF/2013, respectively, and among them 344 nucleotides are identical. Therefore, the two Asian sub-lineages could have diverged from a common ancestor, arrived in Malaysia, established niches, and later dispersed to South America. It is unclear why the ZIKV strain that already existed in 1966 in Malaysia did not have a significant clinical impact until 50 years later in Oceania. A more rigorous analysis of the potential relationship between the genetic changes and epidemiological topography is required and will be possible as

we gain further sequence information on currently circulating clinical strains and their associated

pathology.

Figure 2.1.1 List of Zika virus strains

Name	GenBank No.	Country	Year	Source
MR766-NIID	LC002520.1	Uganda	1947	Monkey (*Macaca mulatta*)
P6-740	HQ234499.1	Malaysia	1966	Mosquito (*Aedes aegypti*)
ArD7117	KF383116.1	Senegal	1968	Mosquito (*Aedes luteocephalus*)
IbH_30656	HQ234500.1	Nigeria	1968	Human blood
ArB1362	KF383115.1	Central African Republic	1968	Mosquito (*Aedes africanus*)
ARB13565	KF268948.1	Central African Republic	1976	Mosquito (*Aedes africanus*)
ARB7701	KF268950.1	Central African Republic	1976-80	Mosquito (*Aedes africanus*)
ARB15076	KF268949.1	Central African Republic	1976-80	Mosquito (*Aedes opok*)
ArD_41519	HQ234501.1	Senegal	1984	Mosquito (*Aedes africanus*)
ArD128000	KF383117.1	Senegal	1997	Mosquito (*Aedes luteocephalus*)
ArD158084	KF383119.1	Senegal	2001	Mosquito (*Aedes dalzieli*)
ArD157995	KF383118.1	Senegal	2001	Mosquito (*Aedes dalzieli*)
FSM	EU545988.1	Micronesia	2007	Human
FSS13025	JN860885.1	Cambodia	2010	Human
PHL/2012/CPC-0740-Asian	KU681082.3	Philippines	2012	Human
H/PF/2013	KJ776791.1	French Polynesia	2013	Human
PLCal_ZV	KF993678.1	Thailand	2013	Human (urine and nasopharynx)
Haiti/1225/2014	KU509998.1	Haiti	2014	Human (plasma)
SV0127_14_Asian	KU681081.3	Thailand	2014	Human
Natal_RGN_Asian	KU527068	Brazil	2015	Human (fetal brain)
Brazil_ZKV2015_Asian	KU497555.1	Brazil	2015	Human (amniotic fluid)
ZikaSPH2015	KU321639.1	Brazil	2015	Human
BeH815744	KU365780.1	Brazil	2015	Human
BeH818995	KU365777.1	Brazil	2015	Human
BeH819015	KU365778.1	Brazil	2015	Human
BeH819966	KU365779.1	Brazil	2015	Human
BeH823339	KU729217.2	Brazil	2015	Human
BeH828305	KU729218.1	Brazil	2015	Human
SSABR1-Asian	KU707826.1	Brazil	2015	Human (serum)
FLR	KU820897.1	Colombia	2015	Human (serum)
103344	KU501216	Guatemala	2015	Human
8375	KU501217.1	Guatemala	2015	Human
PRVABC59	KU501215.1	Puerto Rico	2015	Human
Z1106033	KU312312.1	Suriname	2015	Human (serum)
MRS_OPY_Martinique	KU647676.1	Martinique	2015	Human
VE_Ganxian_Asian	KU744693	Venezuela	2016	Human (serum)
GD01_Asian	KU740184.1	China (imported)	2016	Human
GDZ16001	KU761564.1	Venezuela	2016	Human (saliva)
ZJ03	KU820899.2	China (imported)	2016	Human
Rio-U1	KU926309.1	Brazil	2016	Human (urine, pregnant)
Rio-S1	KU926310.1	Brazil	2016	Human (saliva)
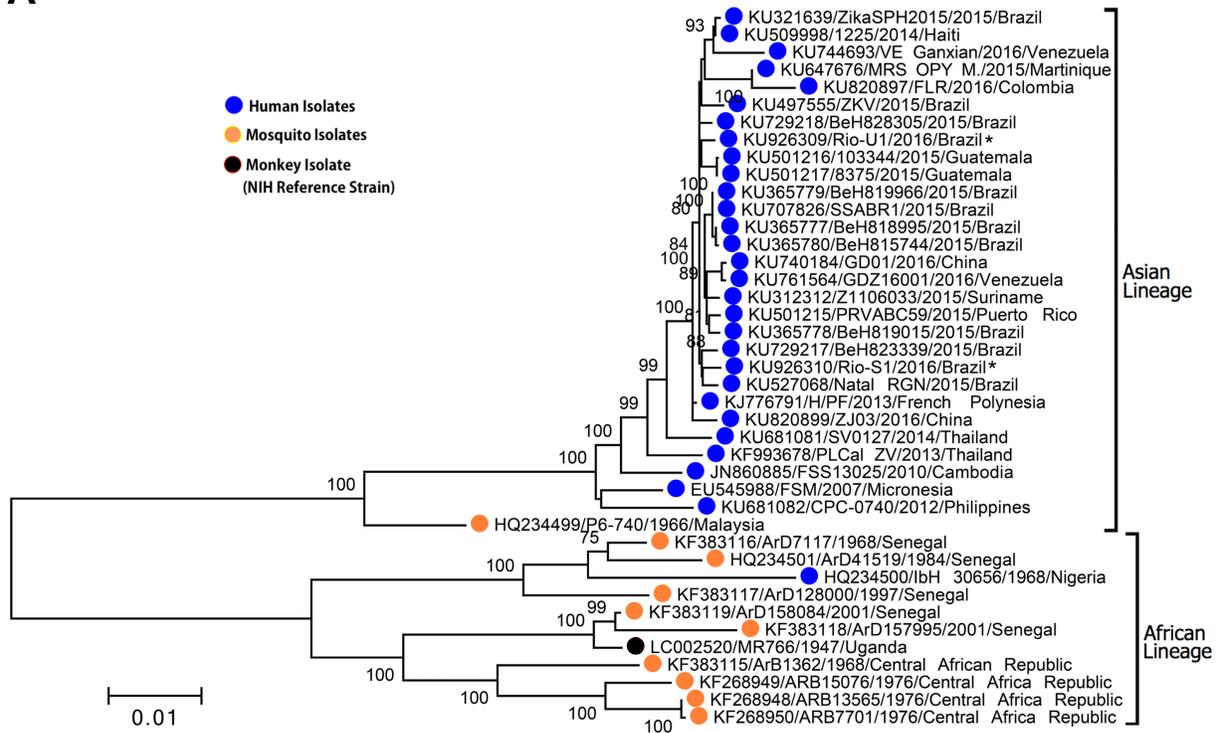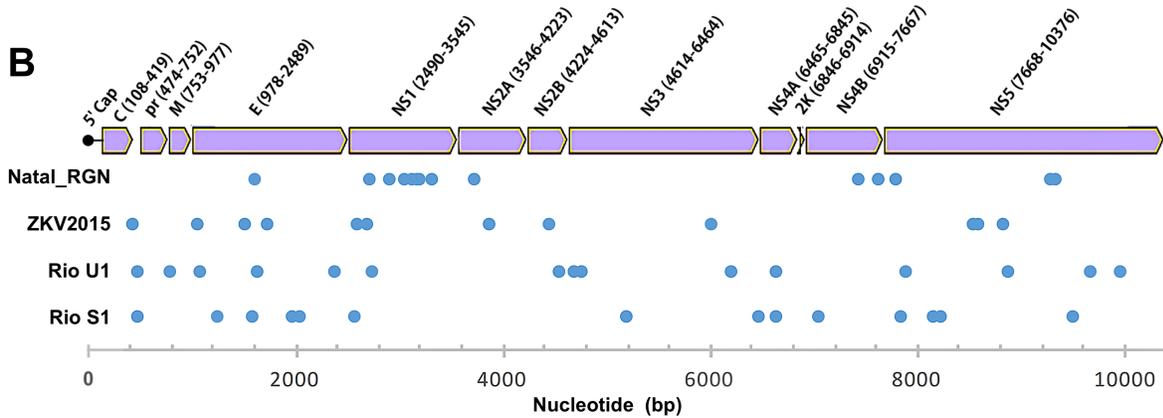
Figure 2.1.2. Evolutionary relationships of Zika virus. All full-length Zika virus (ZIKV) open reading frame nucleotide and amino acid sequences (at least 10,100 bp) were obtained from the NIAID Virus Pathogen Database and Analysis Resource (ViPR) (http://www.viprbrc.org/) and NCBI GenBank. (A) Phylogenetic tree constructed from nucleotide data from 41 viral complete open reading frame sequence of ZIKV strains by the maximum likelihood method logarithm in MEGA7 based on the Tamura-Nei model. A bootstrap percentage for 1000 replicates were shown on the left. Branches corresponding to partitions reproduced in less than 70% bootstrap replicates are not shown. Strains isolated from human, mosquito and monkey (NIH reference strain) were labeled with blue, orange, and black circles respectively. The two subtypes were labeled on the right side of the tree. The new strains Rio-U1 and Rio-S1 were highlighted using (*). (B) Nucleotide alignments were made using MUSCLE from ViPR. Alignment comparisons were

made using Jalview V2.9. Graphical representation of unique nucleotide mutations (blue circle) in Natal_RGN, ZKV2015, Rio-U1, and Rio-S1 strains amongst all (29 total) current human strains within Asian lineage by using pairwise comparisons.
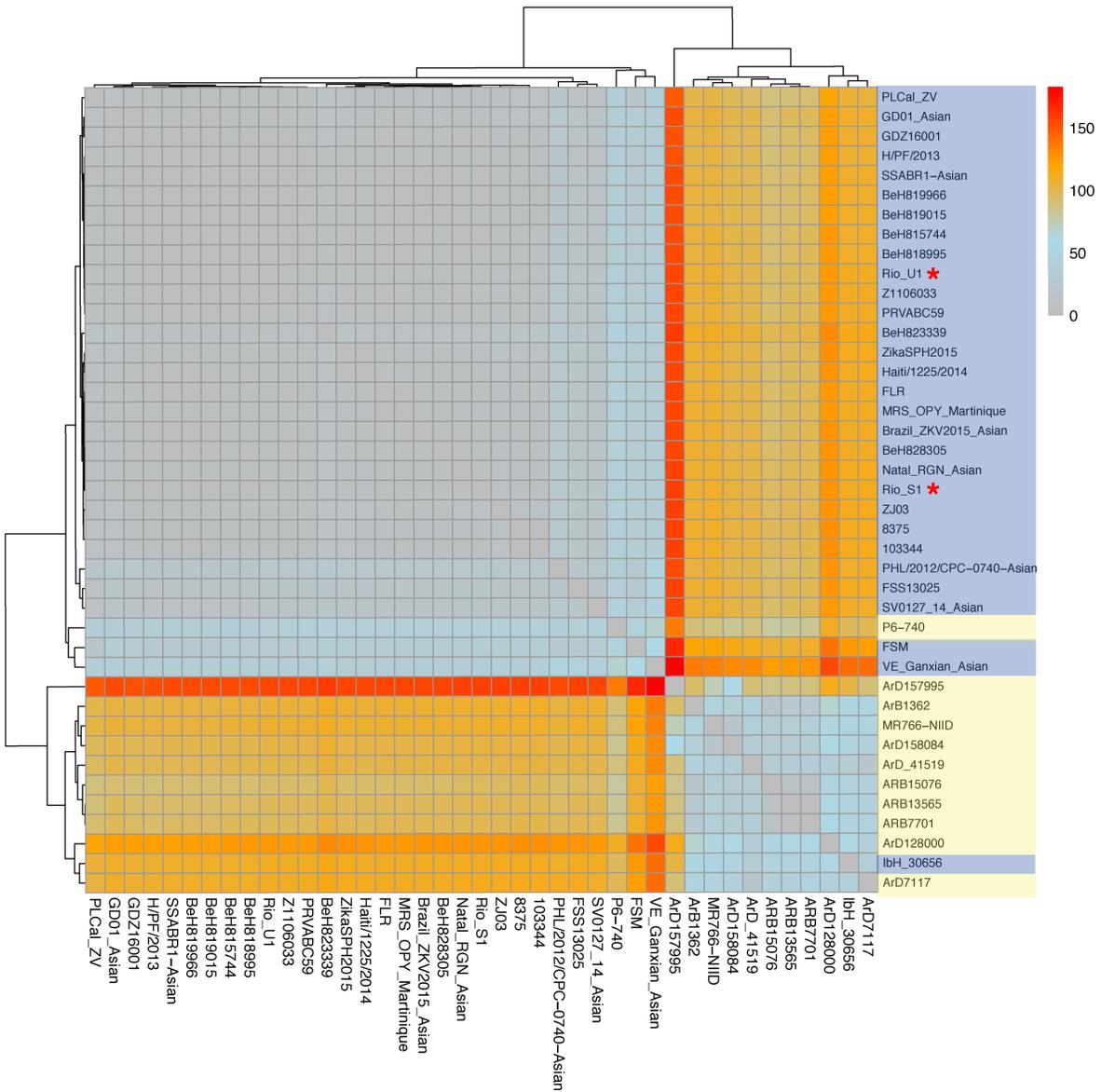
Figure 2.1.3. The heatmap for the amino acid variations of 41 ZIKV strains. Based on the variation table, a heatmap was generated with pheatmap package in R. All viral strains were clustered as two groups with the average geometry distance method. All viral strains were clustered into two lineages based on the number of amino acid variations that indicated as color bars from 0 to 200. The strain names were highlighted as yellow for human and light purple for mosquito-borne Zika viruses.

We focused further analysis on the nucleotide sequence of four independent human strains with known clinical outcomes. Natal_RGN (KU527068) was isolated from the brain tissue of a fetus with severe microcephaly, and ZKV2015 (KU497555.1) was isolated from the amniotic fluid of a pregnant patient whose fetus was diagnosed with microcephaly (Calvet et al., 2016). Here, we include two very recent human isolates: Rio-U1 (KU926309.1) and Rio-S1 (KU926310.1) (bioRxiv doi: http://dx.doi.org/10.1101/045443). Rio-U1 was isolated from the urine of a pregnant woman who presented at 18-weeks gestation with rash and hand arthralgia, edema, and paresthesias. She recovered acutely from her symptoms, her ultrasound at the time of diagnosis was normal, and she continues to be followed. Rio-S1 was isolated from a man who presented with low grade fever, malaise, rash, conjunctival hyperemia, and hand and wrist arthralgias, edema, and paresthesias. His illness self-resolved in 10 days. Compared with the 29 other human strains, the number and distribution of unique nucleotide changes is shown in **Figure 1B.** There were 15, 13, 16, and 15 nucleotide changes in Natal_RGN, ZKV2015, Rio-U1, and Rio-S1, respectively. In pairwise comparisons, there were only three unique amino acid substitutions in Natal_RGN: K940E and T1027A in NS1, and T2509I in NS4B. ZKV2015 had three amino acid substitutions: S550T in E, L1259F in NS2A, and E2831V in NS5. Rio-U1 had only one change: K2039R in NS3. Rio-S1 had three amino acid changes: T625A in E, A2122T in NS4A, V2688A in NS5.

A

B

Asian Lineage

C

	110	143	148	153	157	158		123	139	153	
African Lineage	I	K	A	V	H	V					
Asian Lineage	V	E	P	M	Y	I		V	S	V	**Mosquito Isolate**
								A	N	M	**Human Isolate**

ARB13565/Central Africa Republic/Mosquito

Rio-U1/Brazil/Human

17

Figure 2.1.4: Genetic evolution of the Asian lineage and structural changes in pre-membrane protein (prM). Graphical illustration of comparison between of Zika virus (ZIKV) from both linages. Amino acid comparisons were made between African (8 mosquito strains and 1 monkey strain) and Asian lineages (25 human strains). The sequences are aligned using MUSCLE. Conserved mutations were selected using meta-CATS from ViPR. Graphical map (top bar) of conserved amino acids, shown in blue line. These sites were in the table below, with the addition of an African human isolate (IbH-30656) and an Asian mosquito isolate (P6-740) for reference. Conserved sites from African and Asian lineages were highlighted in orange and blue, whereas non-related substitutions were highlighted in yellow. **(B)** Amino acid substitutions were made from pairwise comparisons of P6-740 against all the Asian lineage isolated from humans. All mutations in P6-740 were labeled in orange. Using P6-740 as a reference, identical amino acid substitutions in the human isolates were also highlighted orange, and differences labeled in blue. **(C)** PrM protein of ZIKV shows significant structural alterations. Amino acid substitutions between strains are shown in the table inset. The cartoon representation of the predicted overall tertiary structural comparison of ZIKV PrM proteins from Rio-U1 (cyan) and ARB13565 (yellow). The automated server program CPHmodels-3.0 was employed to build the model according to the homology modeling method. The model was submitted to the Swiss-Model Workspace to obtain the 3D structure. Then verified using PROCHECK and by Verify3D Structure Evaluation Server and QMEAN. The figure was created in PyMol. The structural templates for the PrM protein query sequences were DENV 1 PrM Protein (PDB_ID: 4b03), which shared 48.35% and 50.55% primary sequence identity with the ZIKV PrM proteins from Rio-U1 and ARB13565, respectively. The N and C termini of the structures were labeled with letters. The differences between these two virus strains were shown in sticks.

We were interested in exploring differences in the protein sequence of the African and Asian lineages. Assuming that an African mosquito subtype was the ancestor of the Asian human subtype, we compared the amino acid sequences between eight African strains (seven from mosquitos and one from monkey) and 25 Asian strains isolated from humans. We found that there were 59 amino acid variations located throughout the viral polyprotein sequence that are shared among the individual strains within the African or Asian lineages, but are different between these two major lineages **(Figure 2A).** For comparison, the African human (IbH-30656/Nigeria/1968) and Asian mosquito (P6-740/Malaysia/1966) strains are shown. Our phylogenic analysis builds upon prior studies by the addition of the most recent human strains to the analysis and further supports the existence of two divergent African and Asian lineages. An important limitation in the

analysis of ancestral strain sequences is the potential that these substitutions were adaptations acquired during passages in mouse brains. This is in comparison to modern isolates that were usually sequenced after low passage numbers in monkey or mosquito cell lines. Viral evolution within the murine host is an important question to consider in future experiments, particularly as mouse models are developed to study pathogenesis (Lazear et al., 2016), as well as antiviral and vaccine development. Further studies are needed to elucidate the sequential acquisition of these mutations and their individual contributions to human pathogenesis.

Due to the recent epidemic and technological advances that now allow rapid and full-length sequencing from direct human isolates, the public bank of human ZIKV sequences has increased from eight at the end of 2014, to 30 as of March 2016. We performed a detailed exploration of the evolution of amino acid polymorphisms of the recent human strains, all of the Asian lineage. Our phylogenetic analysis revealed that all contemporary human isolates share a common ancestor with the P6-740 strain isolated from *Aedes* a*egypti* mosquito. Comparison of protein sequences using P6-740 as the Asian reference, showed that FSM had over 400 variations at the nucleotide level and 26 unique substitutions at the protein level **(Figure 2B)**. Interestingly, when we investigated the sequences of selected human strains identified in more recent epidemics (including FSM, H/PF/2013 and Brazilian strains), we found that all of these strains have acquired changes at an additional eight positions, for a total of 34 amino acid changes compared to P6-740 (**Figure 2B)**. Furthermore, all isolates show identical amino acids at these positions, with the exception of position T2634M/V in the NS5 protein.

Although ZIKV is believed to be primarily transmitted through the mosquito vector, it is interesting to note that no known ZIKV mosquito isolate possesses the same nucleotide sequence as the human strains. One possible explanation is sampling bias, where more recent efforts have focused on isolating the virus from infected humans rather than mosquito arbovirus surveillance.

However, it is notable that Duffy *et al* were unable to detect ZIKV in mosquitos despite active surveillance during the Micronesia outbreak (Duffy et al., 2009). Alternatively, it is possible that other routes of transmission, such as sexual transmission, may have a greater contribution to the wide spread of ZIKV in the Americas. Intriguingly, it was recently reported that New World strains of *Aedes aegypti* and *albopictus* are poor transmitters of ZIKV (Chouin-Carneiro et al., 2016). Clearly, more studies are urgently needed on natural vector transmission of ZIKV in Asia and the Americas, as well as the possibility of a more prominent contribution of alternative modes of transmission.

In addition to alterations in protein sequences and structures during ZIKV evolution, nucleotide sequence changes may have an impact on viral genomic stability, replicative efficiency, and thus viral fitness and transmissibility. Strains from the recent epidemic in Brazil showed 14-18 nucleotide mutations compared to the other strains of the Asian lineage isolated from humans. While the nucleotide changes in ZKV2015, Rio-U1, and Rio-S1 are distributed throughout the viral genomic RNA, 50% of the mutations in Natal_RGN, which was isolated from the brain, are located in the NS1 gene. The phenomenon of tissue-specific mutations has been reported for hepatitis C virus, another flavivirus with infectivity to the brain, liver, and blood (Fishman et al., 2008). No samples from the brain were taken from the fetus with ZKV2015, and therefore tissue-specific evolution of ZIKV cannot be definitively supported from the available data. However, as more samples isolated from different compartments with known clinical outcomes become available, additional genetic and biochemical assays to determine the potential impact of these changes on viral pathogenesis will be possible.

The pr region of prM protein had the highest percentage variability between the Asian human and the African mosquito subtypes. Six of the 59 (~10%) amino acid variations between these subtypes, namely I110V, K143E, A148P, V153M, H157Y, and V158I, were in the pr region

of prM. Furthermore, within the Asian lineage, there were three additional changes in human strains compared to the mosquito strain P6-740: V123A, S139N, and V153M. Structural predictions based on the DENV 1 pr protein (PDB_ID: 4b03) showed significant differences between Rio-U1 and ARB13565 (**Figure 2C)**. Our analysis predicts that A148P could possibly play a critical role in mediating a ten-amino-acid structural change from a loop into a continuous beta sheet. This change was only present in human isolates from both lineages, which suggests a potential relevance in human infectivity.

PrM forms a heterodimer with the main viral surface protein, E, in the neutral pH of the lumen of the endoplasmic reticulum (ER). Immature viral particles translocate from the ER to the highly acidic environment of trans-Golgi network where they are packaged into exosomes. The process of viral maturation takes place in the low pH environment where viral surface proteins go through a drastic conformational rearrangement, due to dissociation of prM-E, formation of E homodimers and exposure of the prM cleavage site to furin protease. The cleaved pr shields the E protein fusion loop throughout the low pH condition to prevent secretion of immature particles from the vesicles and it only dissociates from virions in the extracellular environment (Zhang et al., 2012). The role of prM in viral pathogenesis has been under extensive investigation over the past few years. It has been shown that prM plays a critical role in viral assembly, maturation, heterodimer formation with the E protein, particle secretion, and virulence (Zhang et al., 2012). In our analysis, the six amino acid substitutions in prM (I110V/K143E/A148P/V153M/H157Y/V158I) resulted in a dramatic predicted structural change of prM between the African and Asian strains. The effect of this structural change on viral function is currently unknown, and further investigations are required to determine whether the observed amino acid changes in prM might have altered the viral pathogenesis of the Asian strain.

2.1.c DISCUSSION

It should be noted that Faria *et al*. have published a recent report detailing a similar phylogenetic analysis on sequences obtained from the Brazilian epidemic (Faria et al., 2016). We did not include the sequences from the Faria et al. study since detailed clinical information was not available at the time of our analysis. Further, we utilized an alternative approach to structural modeling—rather than mapping we generated two structural models and overlapped them, leading us to predict the structural change in prM protein.

Our phylogenetic analysis has revealed numerous sequence variations in ZIKV genomes between the African and Asian lineages, as well as among different strains within Asian lineage, as the clinical disease caused by ZIKV has changed from causing only a benign illness to now include severe neuropathology. Our modeling studies suggest that these sequence variations could mediate specific changes in the prM protein, which could play a role in virulence or improved fitness. In addition, we have narrowed these changes to a reasonable number of amino acid or nucleotide changes that can be tested for their effect on ZIKV infectivity. Future experiments will be required to determine which amino acid or nucleotide substitutions are directly responsible for the possible increased neurotropism, heightened viral fitness, and enhanced transmissibility and infectivity from the mosquito vector to the human host.

2.1 REFERENCE

1. Brasil, P., Pereira, J.P., Jr., Raja Gabaglia, C., Damasceno, L., Wakimoto, M., Ribeiro Nogueira, R.M., Carvalho de Sequeira, P., Machado Siqueira, A., Abreu de Carvalho, L.M., Cotrim da Cunha, D*., et al.* (2016). Zika Virus Infection in Pregnant Women in Rio de Janeiro - Preliminary Report. N Engl J Med.

2. Broutet, N., Krauer, F., Riesen, M., Khalakdina, A., Almiron, M., Aldighieri, S., Espinal, M., Low, N., and Dye, C. (2016). Zika Virus as a Cause of Neurologic Disorders. N Engl J Med.

3. Calvet, G., Aguiar, R.S., Melo, A.S., Sampaio, S.A., de Filippis, I., Fabri, A., Araujo, E.S., de Sequeira, P.C., de Mendonca, M.C., de Oliveira, L*., et al.* (2016). Detection and sequencing of Zika virus from amniotic fluid of fetuses with microcephaly in Brazil: a case study. Lancet Infect Dis.

4. Chouin-Carneiro, T., Vega-Rua, A., Vazeille, M., Yebakima, A., Girod, R., Goindin, D., Dupont-Rouzeyrol, M., Lourenco-de-Oliveira, R., and Failloux, A.B. (2016). Differential Susceptibilities of Aedes aegypti and Aedes albopictus from the Americas to Zika Virus. PLoS Negl Trop Dis *10*, e0004543.

5. Duffy, M.R., Chen, T.H., Hancock, W.T., Powers, A.M., Kool, J.L., Lanciotti, R.S., Pretrick, M., Marfel, M., Holzbauer, S., Dubray, C*., et al.* (2009). Zika virus outbreak on Yap Island, Federated States of Micronesia. N Engl J Med *360*, 2536-2543.

6. Faria, N.R., Azevedo, R.d.S.d.S., Kraemer, M.U.G., Souza, R., Cunha, M.S., Hill, S.C., Thézé, J., Bonsall, M.B., Bowden, T.A., Rissanen, I*., et al.* (2016). Zika virus in the Americas: Early epidemiological and genetic findings. Science.

7. Faye, O., Freire, C.C., Iamarino, A., Faye, O., de Oliveira, J.V., Diallo, M., Zanotto, P.M., and Sall, A.A. (2014). Molecular evolution of Zika virus during its emergence in the 20(th) century. PLoS Negl Trop Dis *8*, e2636.

8. Fishman, S.L., Murray, J.M., Eng, F.J., Walewski, J.L., Morgello, S., and Branch, A.D. (2008). Molecular and bioinformatic evidence of hepatitis C virus evolution in brain. J Infect Dis *197*, 597-607.

9. Haddow, A.D., Schuh, A.J., Yasuda, C.Y., Kasper, M.R., Heang, V., Huy, R., Guzman, H., Tesh, R.B., and Weaver, S.C. (2012). Genetic characterization of Zika virus strains: geographic expansion of the Asian lineage. PLoS Negl Trop Dis *6*, e1477.

10. Hills, S.L., Russell, K., Hennessey, M., Williams, C., Oster, A.M., Fischer, M., and Mead, P. (2016). Transmission of Zika Virus Through Sexual Contact with Travelers to Areas of Ongoing Transmission - Continental United States, 2016. MMWR Morb Mortal Wkly Rep *65*, 215-216.

11. Kuno, G., and Chang, G.J. (2007). Full-length sequencing and genomic characterization of Bagaza, Kedougou, and Zika viruses. Arch Virol *152*, 687-696.

12. Lanciotti, R.S., and Lambert, A.J. (2016). Phylogenetic Analysis of Chikungunya Virus Strains Circulating in the Western Hemisphere. Am J Trop Med Hyg.

13. Lazear, Helen M., Govero, J., Smith, Amber M., Platt, Derek J., Fernandez, E., Miner, Jonathan J., and Diamond, Michael S. (2016). A Mouse Model of Zika Virus Pathogenesis. Cell Host & Microbe.

14. Sarno, M., Sacramento, G.A., Khouri, R., do Rosario, M.S., Costa, F., Archanjo, G., Santos, L.A., Nery, N., Jr., Vasilakis, N., Ko, A.I.*, et al.* (2016). Zika Virus Infection and Stillbirths: A Case of Hydrops Fetalis, Hydranencephaly and Fetal Demise. PLoS Negl Trop Dis *10*, e0004517.

15. Zhang, Q., Hunke, C., Yau, Y.H., Seow, V., Lee, S., Tanner, L.B., Guan, X.L., Wenk, M.R., Fibriansah, G., Chew, P.L.*, et al.* (2012). The stem region of premembrane protein plays an important role in the virus surface protein rearrangement during dengue maturation. J Biol Chem *287*, 40525-40534.

2.2 TRACKING THE EVOLUTION OF AND SPREAD OF ZIKA VIRUS BY FOOTPRINT ANALYSIS

ABSTRACT

The recent Zika virus (ZIKV) outbreak in the Americas is believed to have come from Southeast Asia. However, sequencing data to confirm this hypothesis has been limited. Our bioinformatic analyses on all available ZIKV sequences revealed that a strain closely related to one isolated from a Malaysian mosquito in 1966 has continued to circulate in Asia. Genetically diverged Asian strains may have co-evolved in specific patterns over the last half-century before crossing the Pacific. We tracked the evolution of ZIKV through a comprehensive analysis of single nucleotide polymorphisms (SNPs) and classified ZIKV strains in the American lineage into 6 groups, each with a specific signature motif and unique pattern of geographical spread. Using our genotyping method, we characterized a total of 100 ZIKV strains from travelers returning to New York between February and October, 2016, including the first female to male Zika transmission clinical case. This rapid method of differentiating ZIKV viral clades could help with epidemiological tracking of outbreaks and in the future link specific SNPs to clinical outcomes of pathogenicity. In addition to genotyping, we demonstrated the importance of recessive SNPs in NGS data, which should be utilized to aid genome tracing of ZIKV.

2.2.a INTRODUCTION

Zika virus (ZIKV), discovered in 1947, caused only mild and sporadic disease throughout Africa and Asia until the 2007 Micronesia[1] and 2013 French Polynesia outbreaks[2]. The rapid expansion of geographic range[3], the increase in severe pathogenicity including fetal microcephaly[2,4,5], and the novel transmission patterns[6,7] first noted in the 2015-2016 Brazilian epidemic, have raised widespread concern regarding the molecular evolution of this virus[8]. *In vivo* and *in vitro* experiments have confirmed that ZIKV is capable of causing microcephaly and other neurologic damage[9-12]. While the factors behind ZIKV's explosive emergence are still unknown, it has been hypothesized that the virus may have recently evolved to become more neurotropic[13,14], to exhibit increased replicative capacity, and/or to become more transmissible to humans. However, in order to distinguish between the relative contributions of these potential mechanisms, and to help prepare for future outbreaks, we need to have an accurate understanding of ZIKV transmission patterns.

The phylogenetic relationships between ZIKV strains are pivotal to identifying their routes of circulation and evolution. The genetic relationships between the Asian (AS) and African (AF) ZIKV lineages have been previously reported[3,9,15-17]. Single nucleotide polymorphisms (SNPs) have been used in genome-wide association studies as markers of human disease or to trace the evolution of human lineages on a population level. Here, we hypothesized that the frequency of mutation in currently circulating ZIKV strains is adequate for using a SNP-typing method[18] to follow the evolutionary trends in the current epidemic. We analyzed 165 AS human and mosquito ZIKV strains from November 2013 to April 2017, including 3 newly sequenced strains, and generated a set of primers which can be used to differentiate ZIKV groups and lineages, potentially offering a quick and inexpensive method for ZIKV sample analysis.

New York has recorded the largest number of travel-associated Zika virus disease cases in the continental United States (U.S.), with the majority having occurred in New York City (NYC). NYC Department of Health and Mental Hygiene (DOHMH) began developing and implementing plans for managing ZIKV since November 2015. Thus far, we have genotyped 100 samples from travelers in New York from February to October, 2016. Our analysis showed that the majority of the samples are phylogenetically linked to the genotype that circulated in the Caribbean. Our analysis confirmed that several divergent ZIKV groups were concurrently present in the Caribbean islands, providing valuable insight regarding the importation of ZIKV to a major city during a pandemic outbreak. In addition to genotyping, we report epidemiological findings obtained using a new complete ZIKV genomes from the only female to male Zika transmission clinical case seen collected in NYC on July, 2016.

2.2.b RESULTS

Phylogenic analysis reveals prolonged ZIKV transmission in Southeast Asia.

Our previous phylogenetic analysis suggested that the ZIKV strains from the recent outbreak in the Americas were more closely linked with the 1966 Malaysia strain than the strains in the AF lineage. However, it was unclear how a strain that produces only a mild infection evolved as it crossed the Pacific to become so highly pathogenic. Our phylogenetic analysis of the complete genomes of over 500 ZIKV strains to date revealed three major lineages, which can be classified as African (AF), Asian (AS) and American (AM) **(Fig. 1A)**. To characterize key differences between these groups, we compared consensus sequences from each lineage and found that while most amino acid residues are conserved between AS and AM strains, three positions– 139, 2086, and 2634 – were not (**Fig. 1B**). To further characterize AS strains, we constructed the phylogenetic trees based on the ZIKV prM and E genes (**Figs 1C and S1**). We found that the recently isolated sequences for strains from Singapore (ZKA-16–097 and ZKA-16-291), Thailand (BKK01 and SI_BKK01), and China (ZK_YN001) stemmed from a common AS ancestral branch, distinct from the recent AM strains **(Fig. 1C)**. Surprisingly, we found that a strain isolated from a Russian traveler to India (D305/2016/Russia) shared identical amino acid sequence and 98% nucleotide sequence homology in the PrM protein, including the V153 mutation, with the P6_740 strain isolated from a Malaysian mosquito in 1966 **(Fig. 1D)**. These findings suggest that the P6_740 strain may have continued to circulate in Asia over the last half-century.

A

Strains in the recent outbreak

S139N/K
Y2086H/R
M2634V/M

H/PF/2013/French_Polynesia

Strains in Asia

P6-740/1966/Malaysia

Strains in Africa

Asian lineage

African lineage

C

Strains in the recent outbreak

H/PF/2013/French_Polynesia

Strains isolated from Singapore recently

Zika_virus/H.sapiens_tc/THA/2014/SV0127__14l2014_07_19lThailand
BKK01l2016_08_30lThailand
SI_BKK01l2016_08_30lThailand
PLCal_ZVl2013_02_19lCanada
ZIKV/Homo_sapiens/THA/PLCal_ZV/2013l2013_02_12lThailand
ZIKV/Thailand/1601Twl2016_01lThailand
SK403/13ASl2013_09_21lThailand
ZK_YN001l2016_11_03lChina
SV0010/15l2015_01_16lThailand
CTG#0108104l2014_09_01lBangladesh
Zika_virus/H.sapiens_wt/BRA/2016/FC_5905_SERl2016_03_23lBrazil
SK364/13ASl2013_07_09lThailand
Zika_virus/H.sapiens_tc/KHM/2010/FSS13025l2010lCambodia
FSS13025l2010lCambodia
Zika_virus/H.sapiens_tc/PHL/2012/CPC_0740l2012_05_09lPhilippines
ZIKV/Homo_sapiens/PHL/2016/ZK16_0128l2016_09_01lPhilippines
FSMl2007_06lMicronesia
P6_740l1966_07lMalaysia
ZIKV/Aedes_aegypti/MYS/P6_740/1966l1966_07_14lMalaysia
P6_740l1966lMalaysia
ZIKV/Aedes_sp./MYS/P6_740/1966l1966_07_14lMalaysia
D304 l 2016_05_30 l Russia

78 39 79 87 84 31 79 77

B

	C						prM							E											NS1				
	25	27	101	106	110	113	125	139	143	148	153	157	158	246	260	262	410	459	575	607	683	727	728	763	777	785	846	863	988	1030	1058
Africa	N	L	R	A	I	I	I	S	K	A	V	H	V	K	V	V	T	V	F	V	D	V	F	V	T	M	E	K	A	V	V
Asia	S	F	K	T/A	V	V	V	S	E	P	M	Y	I	R	A	A	A	I	S	I	E	A	L	M	M	L	D	R	V	I	M
America	S	F	K	A	V	V	V	N	E	P	M	Y	I	R	A	A	A	I	S	I	E	A	L	M	M	L	D	R	V	I	M

	NS2A						NS2B		NS3										NS4A			NS4B							
	1180	1204	1226	1275	1289	1297	1354	1461	1477	1558	1594	1671	1717	1902	1909	1974	2085	2086	2123	2280	2282	2283	2293	2295	2318	2367	2449	2453	2455
Africa	M	V	V	A	A	I	E	A	A	L	K	K	N	I	M	K	Y	L	N	I	A	A/T/V	M	L	M	I	V	L
Asia	I	A	I	I	V	T	L	D	T	S	H	R	T	H	V	L	R	Y/H	F	S	L	S	A	I	F	I	V	I	S
America	I	A	I	I	V	T	L	D	T	S	H	R	T	H	V	L	R	H	F	S	L	S	A	I	F	I	V	I	S

	NS5																												
	2594	2598	2621	2634	2679	2715	2722	2749	2787	2800	2807	2842	2896	2909	2969	3039	3044	3046	3050	3065	3080	3089	3107	3161	3162	3167	3223	3304	3333	3387
Africa	H	V	R	M	T	M	H	I	A	R	N	V	N	R	H	I	N	A	K	K	E	V	G	K	P	R	S	A	M	D
Asia	Y	I	K	M	A	L	Y	T	V	N	S	I	S	H	Q	V	S	I	R	R	K	I	K	R	S	N	D	S	V	N
America	Y	I	K	V/M	A	L	Y	T	V	N	S	I	S	H	Q	V	S	I	R	R	K	I	K	R	S	N	D	S	V	N

D

Strains in Asia

	125	130	139	143	148	151	153	157	158	246	260	262	263
Strains in Africa	I	S	S	K	A	L	V	H	V	K	V	V	A
P6_740/1966/Malaysia	V	S	S	E	P	L	V	Y	I	R	A	A	A
D304/2016/Russia	V	S	S	E	P	L	V	Y	I	R	A	A	A
FSM/2007/Micronesia	V	S	S	E	P	L	M	Y	I	R	A	A	A
CPC_0740/2012/Philippines	V	S	S	E	P	L	M	Y	I	R	A	A	V
FSS13025/2010/Cambodia	V	N	S	E	P	M	M	Y	I	R	A	A	A
KHM/2010/Cambodia	V	N	S	E	P	M	M	Y	I	R	A	A	A
ZK_YN001/2016/China	V	S	S	E	P	L	M	Y	I	R	A	A	A
THA/2013/Thailand	V	S	S	E	P	L	M	Y	I	R	A	A	A
PLCal_ZV/2013/Canada	V	S	S	E	P	L	M	Y	I	R	A	A	A
SI_BKK01/2016/Thailand	V	S	S	E	P	L	M	Y	I	R	A	A	A
BKK01/2016/Thailand	V	S	S	E	P	L	M	Y	I	R	A	A	A
Tha/2014/Thailand	V	S	S	E	P	L	M	Y	I	R	A	A	A
ZKA_16_291/2016/Singapore	V	S	S	E	P	L	M	Y	I	R	A	A	A
ZKA_16_097/2016/Singapore	V	S	S	E	P	L	M	Y	I	R	A	A	A
NIID123/2016/Japan	V	S	S	E	P	L	M	H	I	R	A	A	A
Other Strains in Singapore	V	S	S	E	P	L	M	Y	I	R	A	A	A
H/PF/2013/French_Polynesia	V	S	N	E	P	L	M	Y	I	R	A	A	A
Strains in the recent outbreak	V	S	N	E	P	L	M	Y	I	R	A	A	A

Figure 2.2.1. Phylogenic analysis reveals prolonged ZIKV transmission in Southeast Asia.
(A) Phylogenetic tree of 432 full-length ZIKV genomes using the GTR model and bootstrap with 100 replicates in PhyML. Strains from the AF, AS and AM lineages are indicated by colored

triangles. Strains from the recent outbreak are highlighted for different sub-clades. Amino acid substitutions are listed on major evolutionary branches. **(B)** Comparison of consensus sequences for AF, AS and AM strains. Highlighted in green are four amino acid residues which were inherited by AF and AS strains, but mutated in AM strains. **(C)** Phylogenetic tree of the prM protein of AS strains. Strains isolated in 2016 are highlighted in red. **(D)** Amino acid substitutions of the ZIKV prM protein as it diverged from AF to AS and later to AM strains.

Comprehensive SNP typing of pandemic ZIKV.

Evolution of ZIKV occurs through the accumulation of mutations that are either inherited from its parent strain or occur *de novo*. From these two distinct sources of mutation, we can infer the evolutionary relationships between ZIKV strains through genotyping (see Material and Methods). To define the "inherited" SNPs (inSNPs) of AM strains, we analyzed the open reading frames (ORFs, nucleotide positions 108-10379) of 500 complete ZIKV genomes, including three newly sequenced strains. InSNPs were recognized if they were present in at least 10 independently isolated strains **(Fig. 2A and S2)**. Within the AM lineage, our genome-wide scan revealed 108 inSNPs, which can be classified based on their genetic homology into the six specific groups shown in the heatmap. Our classification method is consistent with phylogenetic analyses using the GRT model PhyML (**Fig. S3**).

To characterize the variability of ZIKV strains, we analyzed the genome sequences isolated from two groups of Chinese travelers who returned from American Samoa and Venezuela with confirmed ZIKV infection, which were grouped into G1 and G6, respectively **(Fig. 2B and S4)**. We found that the G1 strains matched each other with 99.72% sequence identity **(Table S1)**. In contrast, G6 strains were divided into two subgroups based on two independent sets of shared mutations**.** Interestingly, the travelers from the two G6 subgroups had returned to China from Venezuela during two separate time periods, either at the beginning or end of February 2016. It is therefore possible that these divergent strains were co-circulating in the same region **(Fig. S4)**.

31

These examples further demonstrate the sensitivity and usefulness of SNP-typing for tracking the spread and evolution of ZIKV.

Similar relationships between inSNPs and epidemic patterns were observed in ZIKV strains isolated from Florida, the first state in the U.S. with endemic spread of the virus **(Fig. 2C)**. All strains isolated from humans or mosquitos in Florida could be categorized into group G4, with 31 common inSNPs. The shared inSNPs between human and mosquito strains indicate the local circulation of G4 strains in Miami Dade county. However, our analysis also showed that all G4 strains could be further divided into two sub-groups, which suggests that there may have been more than one event that introduced ZIKV to Florida.

To further characterize the inherited mutations from each group, we analyzed all 108 inSNPs and found 15 amino acid substitutions compared to the reference strain, H/PF/2013 **(Fig. 2D)**. We observed that G1 has the most significant sequence changes compared to H/PF/2013 and the other five groups. We also noted that most mutations occurred in the NS1 and NS5 genes of ZIKV.
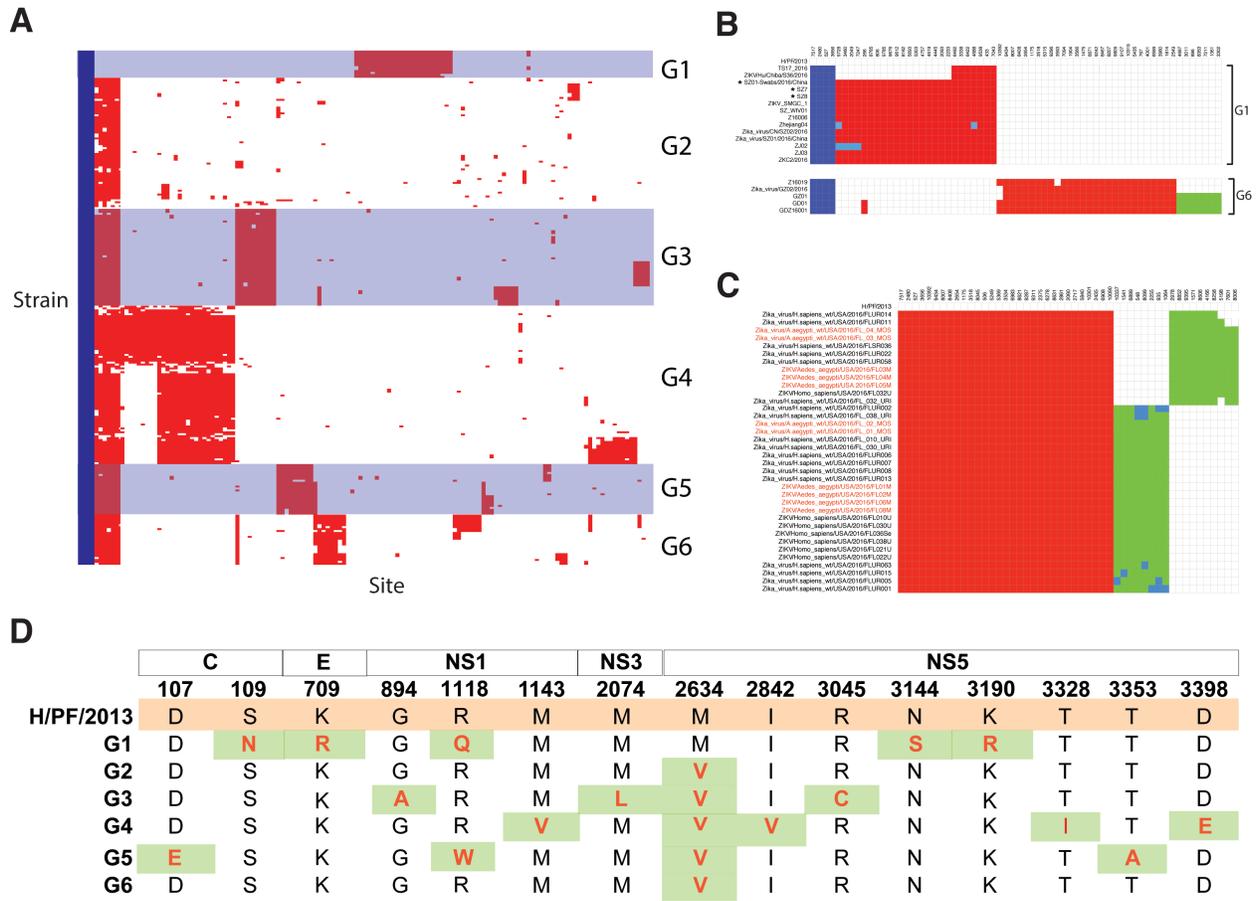
Figure 2.2.2 Comprehensive SNP-typing of pandemic ZIKV. (A) Heatmap for inherited SNPs in ZIKV strains from different groups. **(B)** Heatmap of SNPs for strains isolated from travelers who returned to China during the ZIKV outbreak. **(C)** Group-specific amino acid substitutions using the consensus sequence.

	C		E	NS1			NS3	NS5								
	107	109	709	894	1118	1143	2074	2634	2842	3045	3144	3190	3328	3353	3398
H/PF/2013	D	S	K	G	R	M	M	M	I	R	N	K	T	T	D
G1	D	N	R	G	Q	M	M	M	I	R	S	R	T	T	D
G2	D	S	K	G	R	M	M	V	I	R	N	K	T	T	D
G3	D	S	K	A	R	M	L	V	I	C	N	K	T	T	D
G4	D	S	K	G	R	V	M	V	V	R	N	K	I	T	E
G5	E	S	K	G	W	M	M	V	I	R	N	K	T	A	D
G6	D	S	K	G	R	M	M	V	I	R	N	K	T	T	D

Global and Regional spread of ZIKV

To examine the global spread of ZIKV from Asia to the Americas, we constructed a map depicting the location of infection by all ZIKV strains with complete and partial sequences available as of July 2017 (n=380). **(Fig. 3A)**. We found that the early AM lineage is likely from a common ancestor closely related to the H/PF/2013 strain, which then diverged into six groups that each established local infection in specific geographical regions. G1 has so far only been detected in the Pacific Islands (e.g. Samoa, Fiji). G2 is likely the first to have entered Brazil, and subsequently diverged into G3-6 while spreading across the Americas. Indeed, G2 has been

33

isolated mainly in northeastern Brazil, while G3 is present in Central America (e.g. southern Mexico, El Savador) and G4 mainly circulates in the Caribbean and North America. By contrast, G5 and G6 were found to have a much broader geographical distribution. G0 were isolated in Southeast Asia and had seemly been more similar to ancestral strains previously isolated.

In addition to regions where these strains have established endemic infection, perpetuated by local mosquito populations, different groups of ZIKV have been reported in countries outside of the Americas due to human travel or commercial importation. G1 ZIKV strains normally localized to the Pacific Islands such as Samoa, Fiji and Tonga have been isolated in China, Japan, Vietnam and Australia. G2 ZIKV strains from Brazil and Venezuela have been isolated in Italy and China. G3 ZIKV that normally resides in Central America was also found in the US, including Atlanta, New York and Houston. G4 ZIKV normally found in Puerto Rico has been found in Italy. G6 ZIKV from Venezuela has been found in China. These findings further illustrate the contribution that mosquito carriers and human travel play in amplifying and spreading ZIKV to different parts of the world.
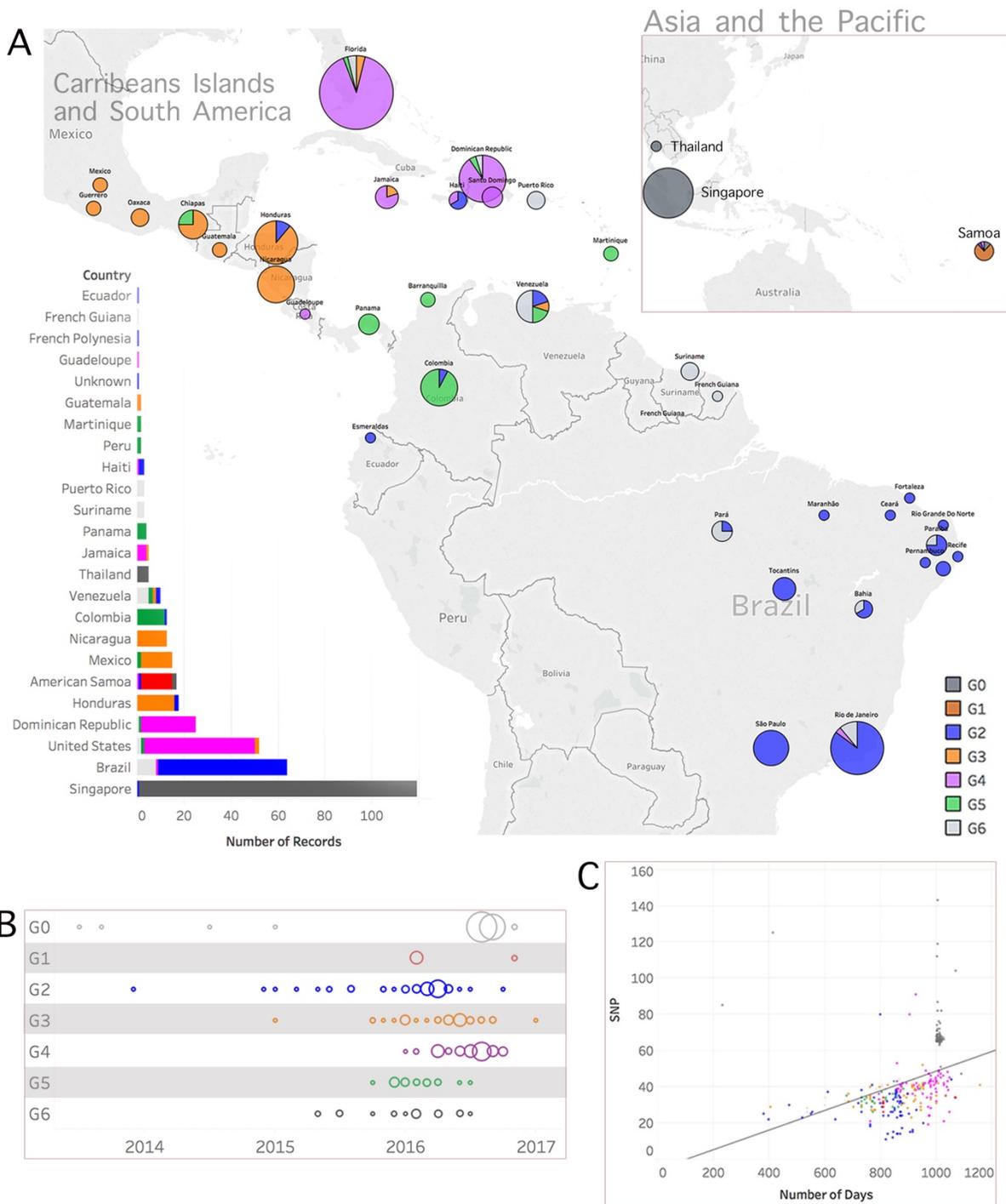
Figure 2.2.3. Global and Regional spread of ZIKV. (A) Geographical mapping of groups 1-6 using the 391 full-length sequences available from GenBank as of July 2017. Strains isolated before H/PF/2013 and the newly isolated strains in Southeast Asia are listed as group 0. The size

35

of the circle indicates the number of strains isolated per location. **(B)** Relationship between the groups and the date of strain isolation. The size of the circle indicates the number of strains isolated per date. **(C)** Rate of ZIKV mutation is extrapolated using the total SNPs from H/PF/2013 and the date of strain isolation.

A rapid method to SNP-type new strains

ZIKV was first reported in the continental United States in July 2016.(23). Nearly 300 local and 5000 imported cases of ZIKV infection have been confirmed as of April 26, 2017. Based on our analysis of the inSNPs from each group, we identified a region of the ZIKV genome (nt 3247-4149) that could be readily amplified and sequenced to classify new strains into our established groups. (Figs. 4A and S6). The relevant inSNP positions for each group are detailed in figure 4A. To test this genotyping method, we successfully sequenced 100 samples obtained from the NYC DOHMH. The original patient population was weighted towards travelers returning from the Dominican Republic (60/100) and Puerto Rico (13/100). Our results confirm that multiple divergent strains are currently circulating in the Caribbean islands, with most samples belonging to G4 (Fig 4B). Samples isolated from travelers returning from South or Central America are consistent with the groups identified in those regions. Interestingly, the three ZIKV-infected travelers from Honduras carried G3 ZIKV.
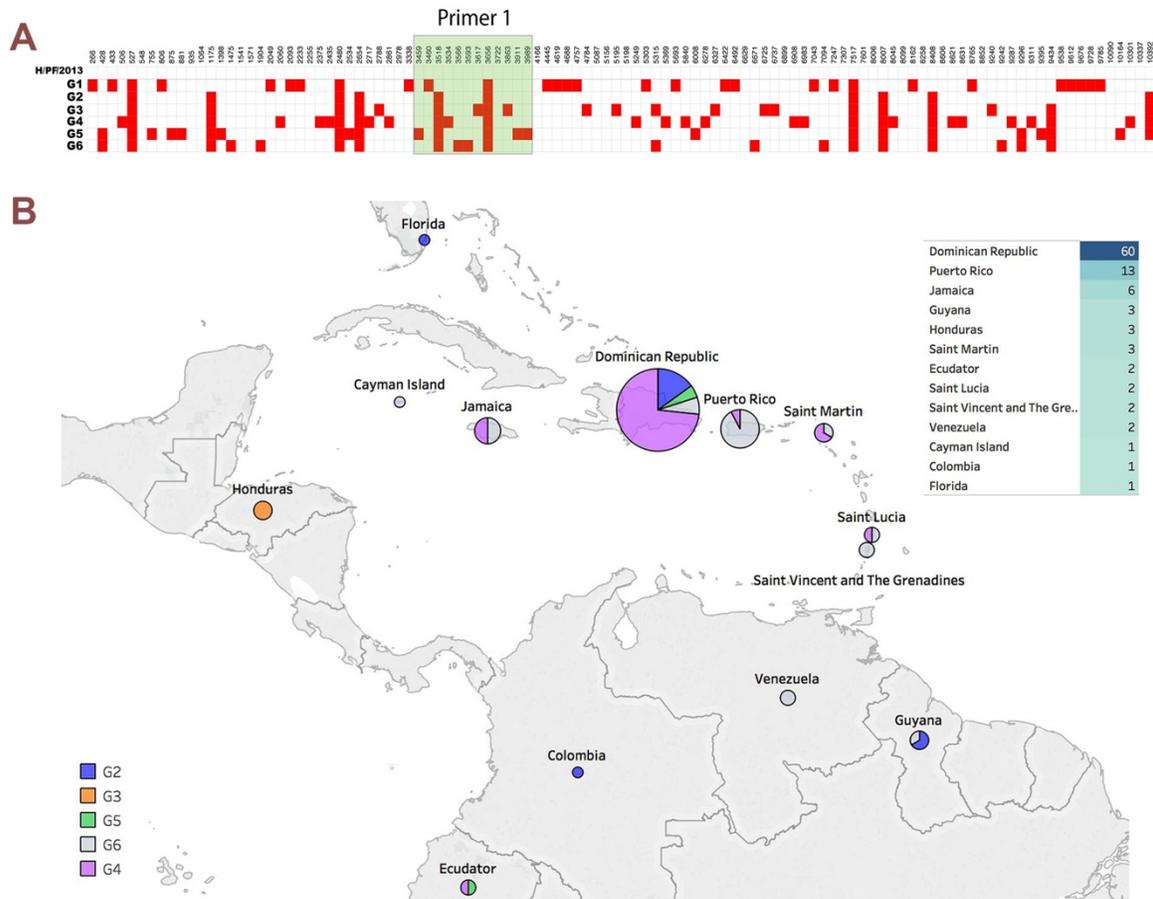
Figure 2.2.4. A rapid method to SNP-type ZIKV samples confirms that divergent strains circulate in the Caribbean. (A) One region of the ZIKV genome was chosen for targeted sequencing. Based on our SNP analysis, this region is sufficient to differentiate all groups based on specific SNPs. (B) Out of 464 PCR-positive samples collected from travelers returning to NYC, we amplified and SNP-typed 100 sequences. Their countries of origin and assigned groups are mapped.

NGS reveals quasi-species evolution in humans and mosquitos

To better understand how a population of ZIKV clones evolve from within a host, we examined the open-sourced NGS database shared by the Anderson Lab(3). We developed a quasi-species distribution analysis pipeline using the consensus sequence for each sample and a >10% variance cut-off. For an inter-host and inter-species comparison of quasi-SNPs, we

37

analyzed both the human (n=10) and mosquito (n=5) ZIKV isolates from Miami, Florida (Fig S7).

We revealed that on average a human sample contains 12.5 quasi-SNPs/sample, which is greater compared to the 4.8 quasi-SNPs/sample seen in the mosquito (Fig 5A). The rates for isolating a unique quasi-SNP were 6.50×10^{-4} and 3.30×10^{-4} in humans and mosquitos, respectively. This is consistent with previous findings suggesting a limited ZIKV replication cycle within the mosquito gut (24, 25). Furthermore, we compared the quasi-SNPs against the full-length ZIKV genome sequences in each of the six groups (Fig 5B). We found two quasi-SNPs, T1064C and T1541G, were unique in the sequences from isolated from the U.S. Interestingly, we found that 4/15 samples contain adenine in position 5315, which can also be found in sequences from G1, 2, 4 and 5. A5315G mutation can be found in G3 and 6 with adenine at position. Samples carrying a mutation at G9344A was only found in G2 and 6.

Quasi-SNP	ZM1	ZM2	ZM3	ZC188	ZF1	ZF8	ZF10	ZF16	ZF21	ZF22	ZF30	ZF32	ZF38	ZF39	ZL2	Number	History NT	G1	G2	G3	G4	G5	G6
1064	T/c	T/c						C/t		C/t			T/c		T/c	6	T/c	T	T	T	T/c	T	T
1541	G/t	G/t					G/t	G/t		G/t					G/t	6	T/g	T	T	T	T/g	T	T
5315				G/a				A/g	G/a	G/a	A/g			A/g	A/g	7	A/g	A	A/g	G	A	A/g	G
9344	G/a	G/a	G/a	A/g			G/a	G/a							G/a	7	G/a	G	G/a	G	G	G	G/a
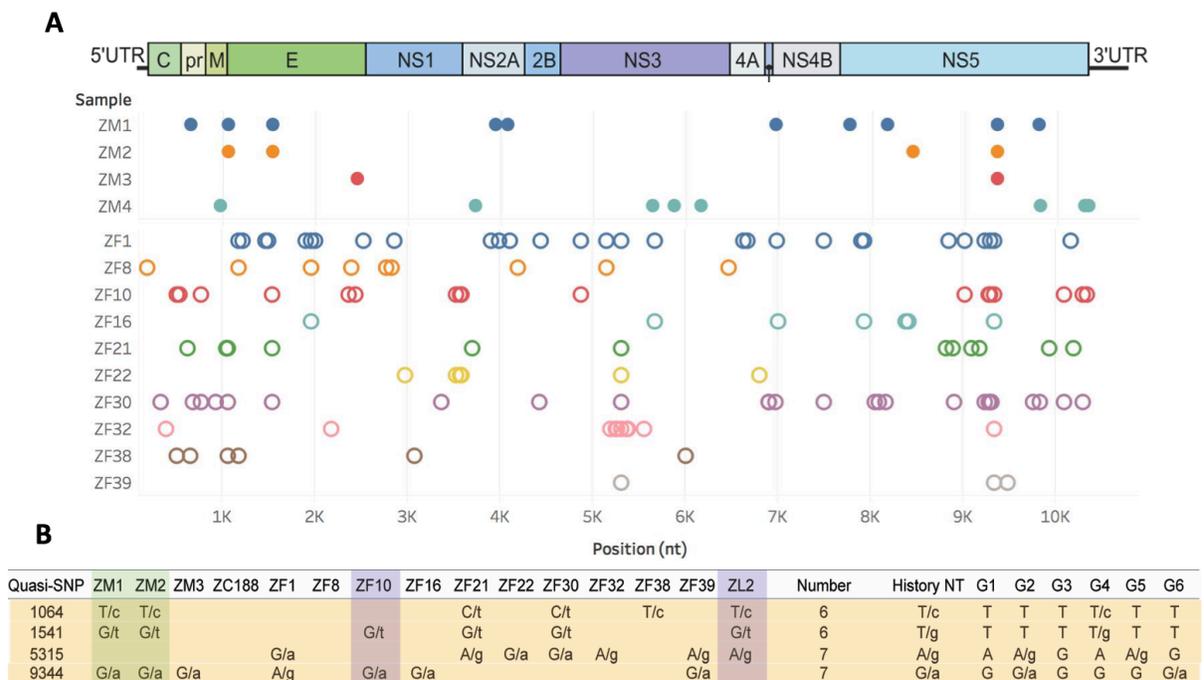
Figure 5. NGS reveals quasi-species evolution in humans and mosquitos. (A) The distribution of quasi-SNPs within 4 mosquito samples (ZM1-ZM4) and 10 human samples (ZF1-ZF39). **(B)** The distribution of four significantly divergent quasi-SNPs (1064, 1541, 5315 and 9344) within groups G1-G6. The major and minor nucleotides of these for sites among 15 samples

(ZM1-ZL2) were distinguished as upper or lower cases, such as T/c. Also, the nucleotide compositions of these four sites within each genotyping groups (G1-G6) were shown.

2.2.c DISCUSSION

In the global effort to combat Zika virus, rapid analysis of the genetic variation and evolutionary relationships of the virus is critical to early epidemiological investigations. We found that all pandemic ZIKV strains carry traceable inSNPs that can be used as molecular footprints. This information allowed us to determine patterns of viral evolution and develop potential tools to rapidly classify unidentified ZIKV strains.

Our previous ZIKV genome sequence analyses revealed that the ZIKV strains in the American outbreak are likely derived from the Asian instead of the African lineage. However, it was unclear how AS strains such as P6_740 could spread to the Americas a half-century later and cause much higher rates of microcephaly. Through a comprehensive analysis of the complete and partial DNA sequences of the ZIKV strains isolated in Asia, we developed an improved understanding of the spread and evolution of the ZIKV AS lineage. First, we found that the ZIKV isolated from a traveler to India in 2016 shared an identical amino acid sequence in the PrM region as the P6_740 strain, suggesting that ZIKV could have been continuously circulating in Asia over the past half-century. Second, ZIKV strains isolated from different countries in Asia are more genetically divergent than the strains found in Americas. While P6_740 is closer to the AF lineage, NIID123/2016/Japan is closer to the AM lineage, suggesting that the original AS strains may have undergone a series of mutations before migrating to the Americas. Third, most of the ZIKV strains in Asia are only mildly symptomatic, which could explain why its continuous circulation has gone unnoticed for decades. However, the recent finding of two cases of ZIKV in Thailand that caused microcephaly indicates that some AS strains of ZIKV may also be evolving to highly pathogenic viruses. Our studies have identified multiple sequence differences between different strains in the AS lineage. It would be interesting to identify mutations that are potentially associated with pathogenicity, although additional factors such as previous exposures to other flaviviruses may contribute to the severity of the disease[20].

We have identified 108 SNPs in the 165 ZIKV strains isolated from the recent outbreaks, and developed a SNP-typing strategy to classify them into 6 distinct groups. This approach has allowed us to propose evolutionary pathways for each group, following assumed patterns of SNP inheritance between parental and progeny strains and accounting for *de novo* mutations that define new groups. This approach could become a valuable complement to traditional phylogenetic analyses that have been extensively generated for ZIKV to infer viral evolution. This method has several merits. First, it presents the similarities and variations between viral strains at the nucleotide position level, unlike RNA viruses such as hepatitis C and influenza, which acquire too many mutations to employ this SNP-typing methodology. Second, it assists in the identification of strain and group-specific modules of variation by SNP markers. However, we should point out that each group may be further assigned into sub-groups due to the accumulation of mutations in the adaptive evolution of viruses (**Fig. S6**). Although our method may be limited by the number of ZIKV samples sequenced to date, we believe that this system has the capacity to be further improved once more viral sequences are obtained.

ZIKV sequences from several groups of Chinese travelers provided us with another opportunity to test our SNP analysis model. We were able to classify the viruses as belonging to either G1 or G6, respectively, based on specific molecular signatures. The G1 strain was isolated from people who traveled to American Samoa and returned to China together. All travelers were found to have very similar ZIKV sequences, with only a few strain-specific mutations, strongly suggesting that they had all been initially infected by the same ZIKV strain circulating in that area. By contrast, G6 ZIKV strains were isolated from patients who returned to China from Venezuela, and we were able to identify two subgroups of the virus that correlated with the likely timing of infection (i.e. whether the travelers had returned at the beginning or the end of February 2016). These examples further demonstrate the sensitivity of our system, which can potentially

identify the geographic region where ZIKV strains originated, and provide clues regarding temporal distribution when multiple strains are in circulation. On a global level, mapping of the complete ZIKV sequence further revealed regional spread of a unified genotype of ZIKV. This finding provides useful information to further investigate the virus's genotypic adaptations to particular regions, and improve our understanding of its spread for early epidemiological control. Although these conclusions are based on the current number of full genome sequences, it is possible that additional genome sequences will reveal a more diverse population of strains.

We describe an easy-to-use method to quickly genotype isolated ZIKV strains using specific primers for Sanger sequencing, without the need for full-genome sequencing. This method can be effectively used in limited settings using only basic equipment, at only a fraction of the cost. The samples collected in NYC, of which over 50% came from travelers to the Dominican Republic, were confirmed by our analysis to have a majority of G4 strains, which are the group predominantly circulating in the Caribbean. We acknowledge that these samples may be biased towards this group, as NYC has a large Dominican immigrant population. However, as more samples are collected in the future, we expect the groupings identified by our system will reflect the patterns of ZIKV spread as it evolves to new areas. Thus, our system can be used to facilitate the interpretation of sequence and surveillance data, and provides a framework from which new evolutionary branches could continue to be defined as our sample size grows.

ZIKV has been particularly alarming due to its ability to transmit sexually. Previous work has demonstrated male to female or male to male transmission (Ref). Here we characterized the complete genome of two patients from the first reported "female to male" transmission of ZIKV. Our data revealed X mutation… Since there are no know active establishment of *Aedes aegypti* in the city of New York, our evidence supports the hypothesis that ZIKV transmission can occur without any gender restriction.

We anticipate that genotyping of ZIKV will become increasingly useful once molecular signatures can be combined with functional models to determine the clinical outcomes associated with each group. Although it is known that NYC has more imported ZIKV cases compared to Florida, and evidently from similar sources, no local transmission there has been reported. This suggest that human-to-human transmission of ZIKV may have a low impact on ZIKV transmission. It is possible that local environmental and host genetic factors have a greater influence on the clinical outcomes observed in these areas. As ZIKV continues to evolve and spread across the globe, it becomes critical that we develop one universal model which can accommodate our growing knowledge of this virus.

MATERIAL AND METHODS

Isolation and sequencing of ZIKV

The ZIKV SZ01/s strain was isolated from the third imported ZIKV case [21] to mainland China. SZ07 (Zika virus/SZ07/2016/China) and SZ08 (Zika virus/SZ08/2016/China) were isolated from the blood of two patients who were in the same tour group as the third imported ZIKV case, and had returned to Zhejiang, China from American Samoa. For isolating and sequencing ZIKV, Vero cells were seeded in 24 well plates for 24 hours before inoculation, then swab and blood samples were diluted in DMEM supplemented with 2% FBS (1:5) and used to inoculate Vero cells. After 1 h incubation at 37°C, the supernatant was replaced with 0.5 mL cell culture medium. Total RNA was extracted from cell supernatants 5 days after inoculation using the PureLink® RNA Mini Kit (Life technologies, USA) and amplified using the SuperScript III One-Step RT-PCR System (Life Technologies) with a total of 15 primer pairs to generate the overlapping amplicons spanning the entire coding sequence (CDS) as described previously[21,22]. All sequencing was carried out using an ABI 3730 Sanger-based genetic analyzer, and the genome was assembled using DNASTAR version 7.0.

ZIKV sequence analysis

In total, 223 full-length (at least 10,100 bp) and 85 partially sequenced Zika virus nucleotide and amino acid sequences published as of March 22, 2017 were collected. Except for three newly reported sequences (SZ01/s, SZ07 and SZ08), the other 220 genomes were obtained from the NIAID Virus Pathogen Database and Analysis Resource (ViPR) (http://www.viprbrc.org/) and NCBI GenBank, including the reference strain (H/PF/2013). The phylogenetic tree was constructed from nucleotide data from 223 complete sequences of ZIKV strains using the GRT model and bootstrap with 100 replicates in PhyML[23]. Alignment comparisons were made using Seaview. The numbering system for genomic nucleotide and amino acid residue locations is based on the genome or polyprotein of the H/PF/2013 strain, respectively.

SNP analysis and Heatmap

Two basic assumptions are made in the ZIKV evolutionary analyses. First, progeny viruses should inherit all mutations from their parental strain. Second, the baseline mutation rate is high enough that progeny viruses can be expected to gain unique mutations, and should therefore carry more mutations than their ancestral strains. Directed by these two principles, we analyzed and defined different SNP-based groups of ZIKV. In total, 165 AS human and mosquito ZIKV strains were aligned and compared against a reference strain H/PF/2013 (Fig. S3). Each site in the comparison was labeled as either "0" or "1" to indicate a matching nucleotide or polymorphism (SNP). To adjust for possible artifacts, SNPs common to multiple strains (≥10 in Fig. 2B and Fig. S2) are displayed. In total, 108 nucleotide locations mutated in at least 10 different strains were identified, from which we created a matrix to compare 165 ZIKV strains. Based on the matrix of 108 inSNPs * 165 strains, two directional clustering was made for both sites and strains via Pheatmap package in R [24].

Grouping of ZIKV strains

All 165 ZIKV strains except the reference strain, H/PF/2013, were separated into six groups according to the shared mutation block in the heatmap, which were later defined as group-specific mutations. For example, eleven strains (TS17_2016, SZ_WIV01, Z16006, Zhejiang04, Zika_virus.CN.SZ02.2016, Zika_virus. SZ01.2016.China, ZIKV.Hu.Chiba.S36.2016, ZIKV_SMGC_1, ZJ02, ZJ03 and ZKC2.2016) have a specific mutation block that includes 10 sites (Fig. S2). Therefore, these 11 strains were labeled as Group G1, and these 10 sites were identified as G1-specific mutations. The other five groups and their specific mutations were also identified similarly (Fig. 2A). If we reduce the threshold to define SNPs as mutations in at least 2 strains, the number mutation will increase dramatically (Fig. S6). Once all strains have been

clustered into groups based on mutations, group-specific features of spatial and temporal distribution can be further investigated (Figs 3 and S7).

Identification of amino acid substitutions

For all 108 inSNPs identified previously, their corresponding 15 amino acid substitutions were confirmed. The substitutions were identified based on the consensus sequence of each group and the reference strain, H/PF/2013. The consensus sequences used in Fig. 1B, 1D and 2D are generated with Simple Consensus Maker (https://www.hiv.lanl.gov/content/sequence/CONSENSUS/SimpCon.html).

REFERENCE

1.	Lanciotti RS, Lambert AJ. Phylogenetic Analysis of Chikungunya Virus Strains Circulating in the Western Hemisphere. Am J Trop Med Hyg. 2016.

2.	Broutet N, Krauer F, Riesen M, Khalakdina A, Almiron M, Aldighieri S, et al. Zika Virus as a Cause of Neurologic Disorders. N Engl J Med. 2016.

3.	Grubaugh ND, Ladner JT, Kraemer MUG, Dudas G, Tan AL, Gangavarapu K, et al. Genomic epidemiology reveals multiple introductions of Zika virus into the United States. Nature. 2017;546(7658):401-5.

4.	Hills SL, Russell K, Hennessey M, Williams C, Oster AM, Fischer M, et al. Transmission of Zika Virus Through Sexual Contact with Travelers to Areas of Ongoing Transmission - Continental United States, 2016. MMWR Morb Mortal Wkly Rep. 2016;65(8):215-6.

5.	WHO. Zika Virus Situation Report (March). 2016.

6.   Schuler-Faccini L, Ribeiro EM, Feitosa IM, Horovitz DD, Cavalcanti DP, Pessoa A, et al. Possible Association Between Zika Virus Infection and Microcephaly - Brazil, 2015. MMWR Morb Mortal Wkly Rep. 2016;65(3):59-62.

7.   Brasil P, Pereira JP, Jr., Raja Gabaglia C, Damasceno L, Wakimoto M, Ribeiro Nogueira RM, et al. Zika Virus Infection in Pregnant Women in Rio de Janeiro - Preliminary Report. N Engl J Med. 2016.

8.   Deckard DT, Chung WM, Brooks JT, Smith JC, Woldai S, Hennessey M, et al. Male-to-Male Sexual Transmission of Zika Virus - Texas, January 2016. MMWR Morb Mortal Wkly Rep. 2016;65(14):372-4.

9.   D'Ortenzio E, Matheron S, de Lamballerie X, Hubert B, Piorkowski G, Maquart M, et al. Evidence of Sexual Transmission of Zika Virus. N Engl J Med. 2016.

10.   Shi W, Zhang Z, Ling C, Carr MJ, Tong Y, Gao GF. Increasing genetic diversity of Zika virus in the Latin American outbreak. Emerg Microbes Infect. 2016;5(7):e68.

11.   Miner JJ, Cao B, Govero J, Smith AM, Fernandez E, Cabrera OH, et al. Zika Virus Infection during Pregnancy in Mice Causes Placental Damage and Fetal Demise. Cell. 2016;165(5):1081-91.

12.   Tang H, Hammack C, Ogden SC, Wen Z, Qian X, Li Y, et al. Zika Virus Infects Human Cortical Neural Progenitors and Attenuates Their Growth. Cell stem cell. 2016;18(5):587-90.

13.   Wu KY, Zuo GL, Li XF, Ye Q, Deng YQ, Huang XY, et al. Vertical transmission of Zika virus targeting the radial glial cells affects cortex development of offspring mice. Cell Res. 2016.

14.   Heymann DL, Hodgson A, Sall AA, Freedman DO, Staples JE, Althabe F, et al. Zika virus and microcephaly: why is this situation a PHEIC? Lancet. 2016;387(10020):719-21.

15.   Marrs C, Olson G, Saade G, Hankins G, Wen T, Patel J, et al. Zika Virus and Pregnancy: A Review of the Literature and Clinical Considerations. Am J Perinatol. 2016;33(7):625-39.

16.   Wang L, Valderramos SG, Wu A, Ouyang S, Li C, Brasil P, et al. From Mosquitos to Humans: Genetic Evolution of Zika Virus. Cell Host Microbe. 2016;19(5):561-5.

17.   Haddow AD, Schuh AJ, Yasuda CY, Kasper MR, Heang V, Huy R, et al. Genetic characterization of Zika virus strains: geographic expansion of the Asian lineage. PLoS Negl Trop Dis. 2012;6(2):e1477.

18.   Brown C. Zika virus outbreaks in Asia and South America. CMAJ. 2016;188(2):E34.

19.   Zhu Z, Chan JF, Tee KM, Choi GK, Lau SK, Woo PC, et al. Comparative genomic analysis of pre-epidemic and epidemic Zika virus strains for virological factors potentially associated with the rapidly expanding epidemic. Emerg Microbes Infect. 2016;5:e22.

20.   Faria NR, Azevedo RdSdS, Kraemer MUG, Souza R, Cunha MS, Hill SC, et al. Zika virus in the Americas: Early epidemiological and genetic findings. Science. 2016.

21.   Baaijens JA, Aabidine AZE, Rivals E, Schonhuth A. De novo assembly of viral quasispecies using overlap graphs. Genome Res. 2017;27(5):835-48.

22.   van Boheemen S, Tas A, Anvar SY, van Grootveld R, Albulescu IC, Bauer MP, et al. Quasispecies composition and evolution of a typical Zika virus clinical isolate from Suriname. Sci Rep. 2017;7(1):2368.

23.   Hennessey M, Fischer M, Staples JE. Zika Virus Spreads to New Areas - Region of the Americas, May 2015-January 2016. MMWR Morb Mortal Wkly Rep. 2016;65(3):55-8.

24.   Ciota AT, Bialosuknia SM, Ehrbar DJ, Kramer LD. Vertical Transmission of Zika Virus by Aedes aegypti and Ae. albopictus Mosquitoes. Emerg Infect Dis. 2017;23(5):880-2.

25.   Thangamani S, Huang J, Hart CE, Guzman H, Tesh RB. Vertical Transmission of Zika Virus in Aedes aegypti Mosquitoes. Am J Trop Med Hyg. 2016;95(5):1169-73.

26.   Dejnirattisai W, Supasa P, Wongwiwat W, Rouvinski A, Barba-Spaeth G, Duangchinda T, et al. Dengue virus sero-cross-reactivity drives antibody-dependent enhancement of infection with zika virus. Nat Immunol. 2016;17(9):1102-8.

27.   Deng YQ, Zhao H, Li XF, Zhang NN, Liu ZY, Jiang T, et al. Isolation, identification and genomic characterization of the Asian lineage Zika virus imported to China. Sci China Life Sci. 2016;59(4):428-30.

28.   Zhang FC, Li XF, Deng YQ, Tong YG, Qin CF. Excretion of infectious Zika virus in urine. Lancet Infect Dis. 2016.

29.   Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 2010;59(3):307-21.

30.   Gouy M, Guindon S, Gascuel O. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol. 2010;27(2):221-4.

31.   Kolde R, Vilo J. GOsummaries: an R Package for Visual Functional Annotation of

Experimental Data. F1000Res. 2015;4:574.

Supplementary Figures
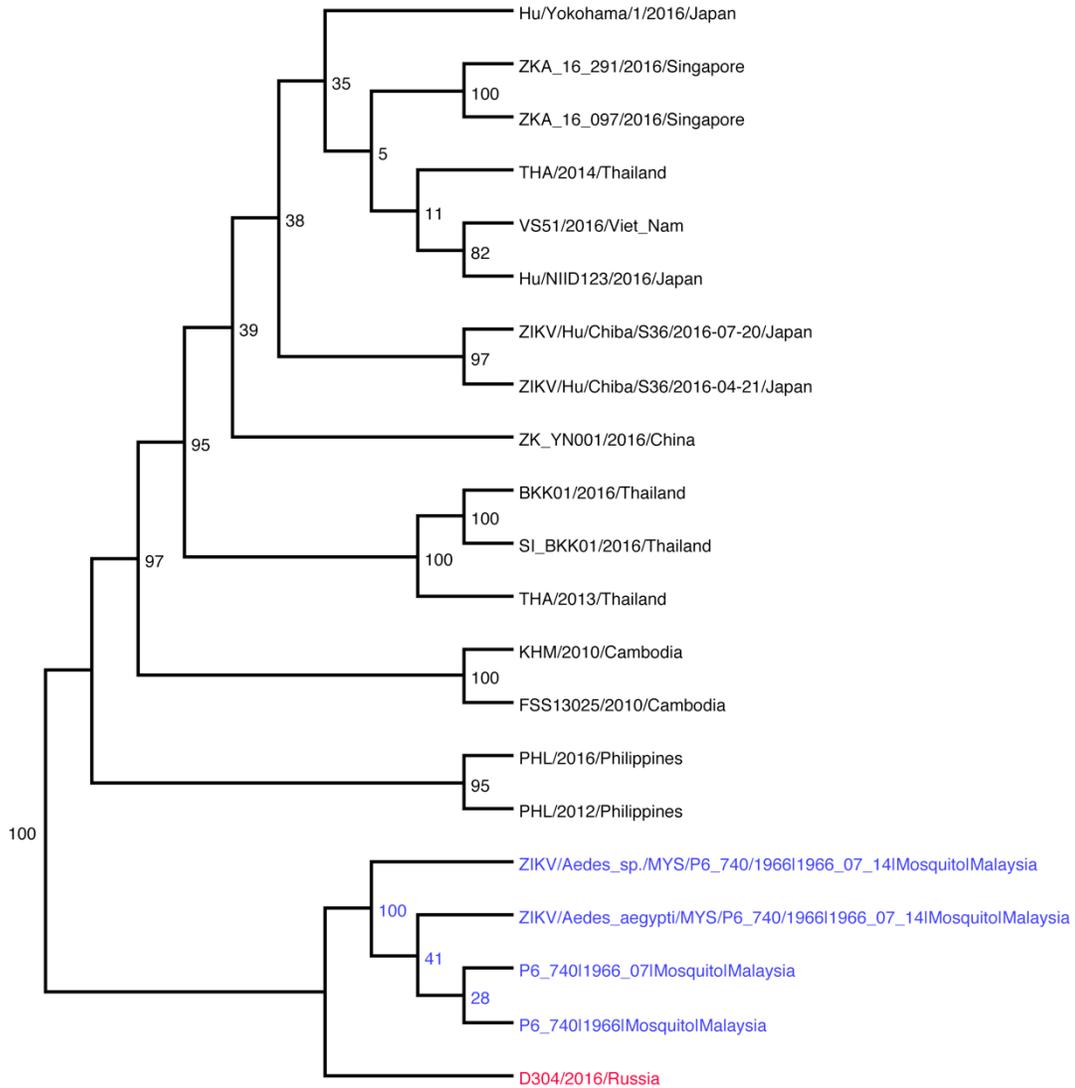
Figure 2.2.S1. Phylogenetic tree of the E gene of Asian strains. Strains D304/2016/Russia and P6_740 are highlighted in red and blue, respectively.

51
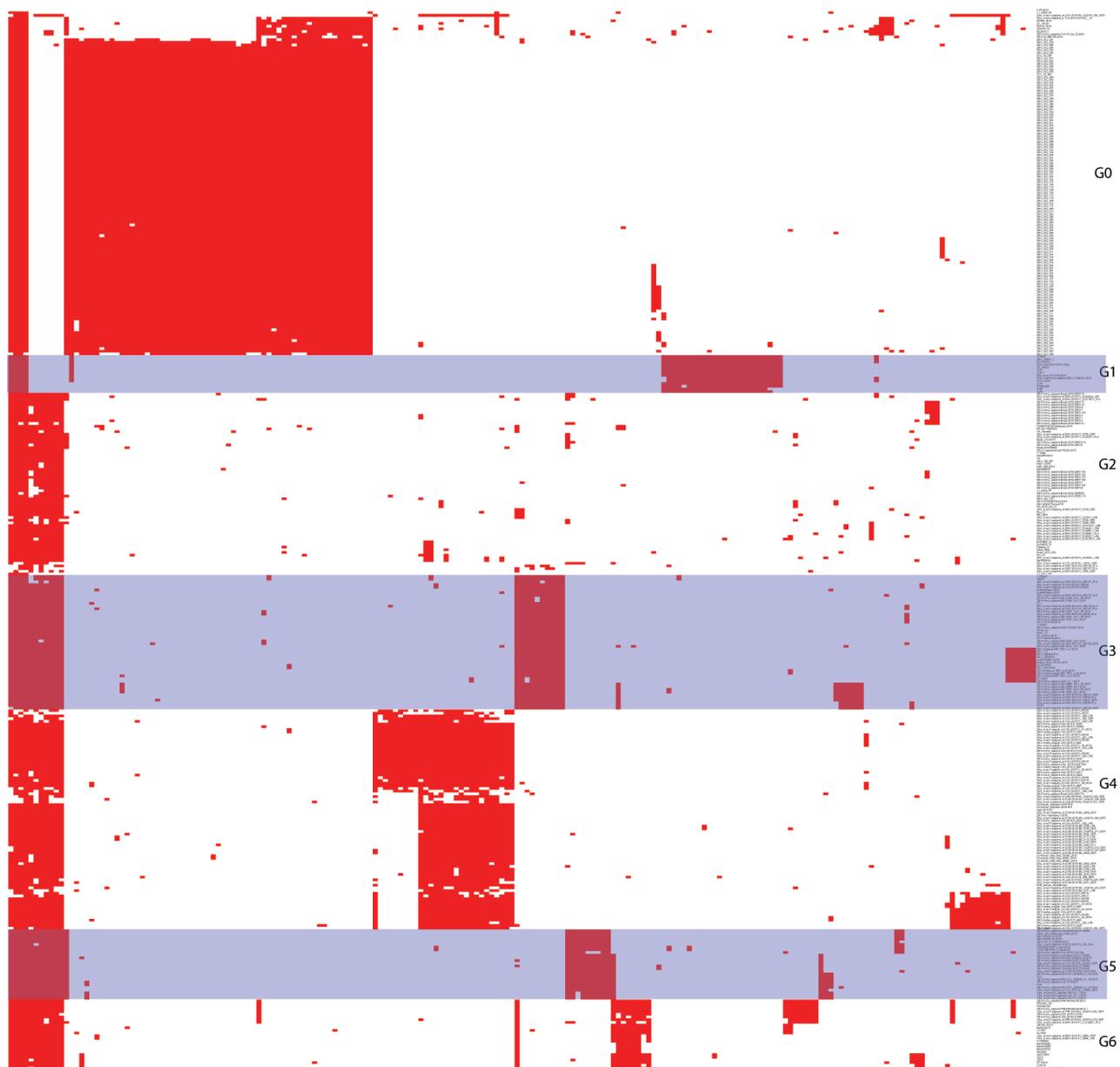
Figure 2.2.S2. Heatmap of 108 identified inherited SNPs from 165 AS human and mosquito ZIKV strains were grouped to G1-G6 based on genetic homology. H/PF/2013 was used as the reference strain (row) and nucleotide sites (column) were grouped with Pheatmap package in R.
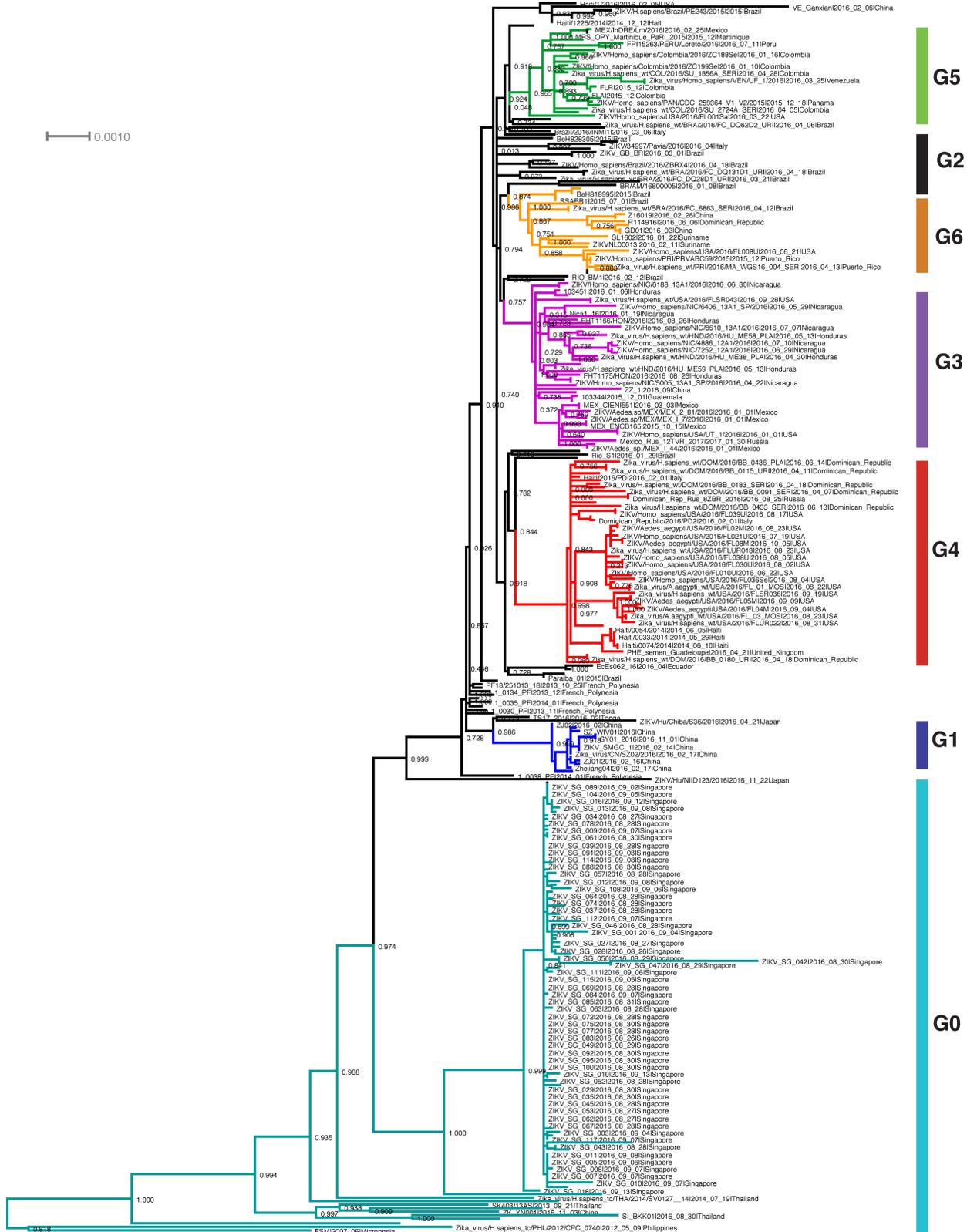
Figure 2.2.S3. Mapping of SNP-based groups using a phylogenetic tree. H/PF/2013 was used as the reference strain. Different groups are shown in different colors.

53
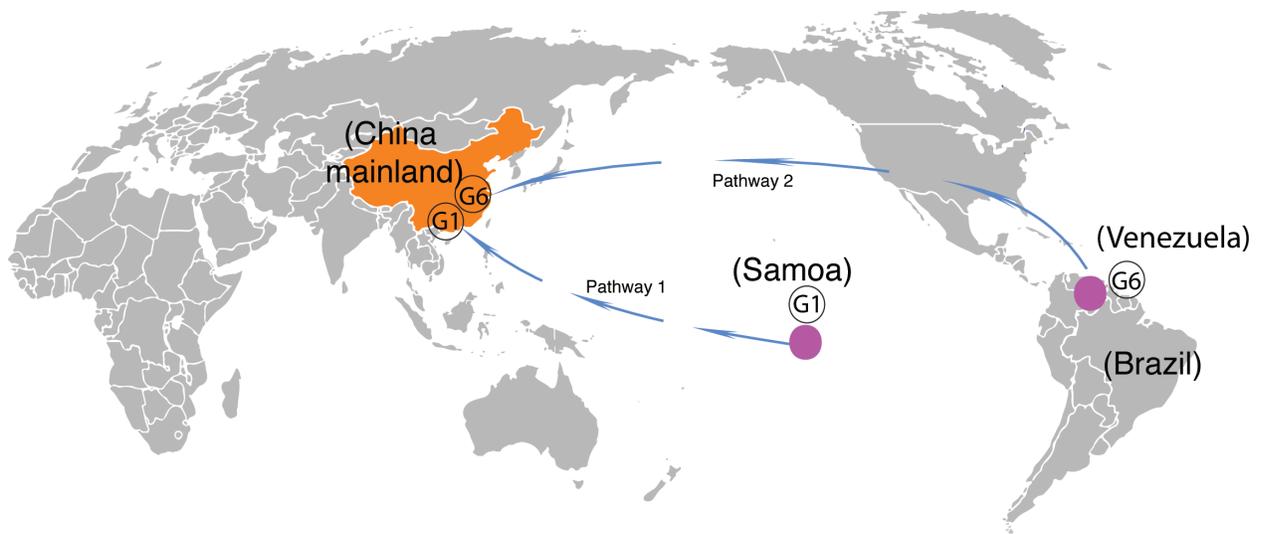
Figure 2.2.S4. Transmission pathways for strains isolated from travelers returning to China during the ZIKV outbreak.
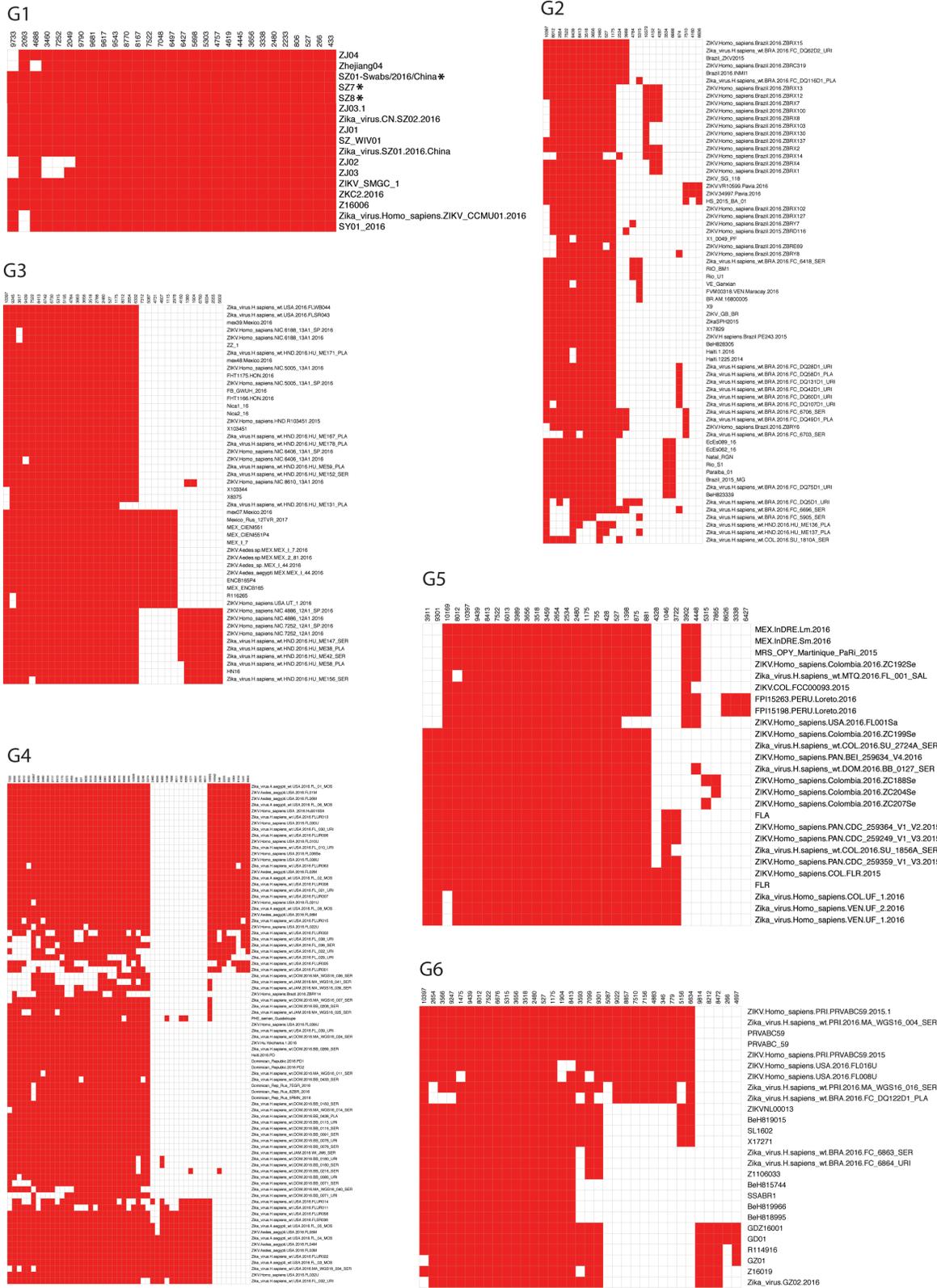
Figure 2.2.S5. Heatmap of inherited SNPs for each group.

55
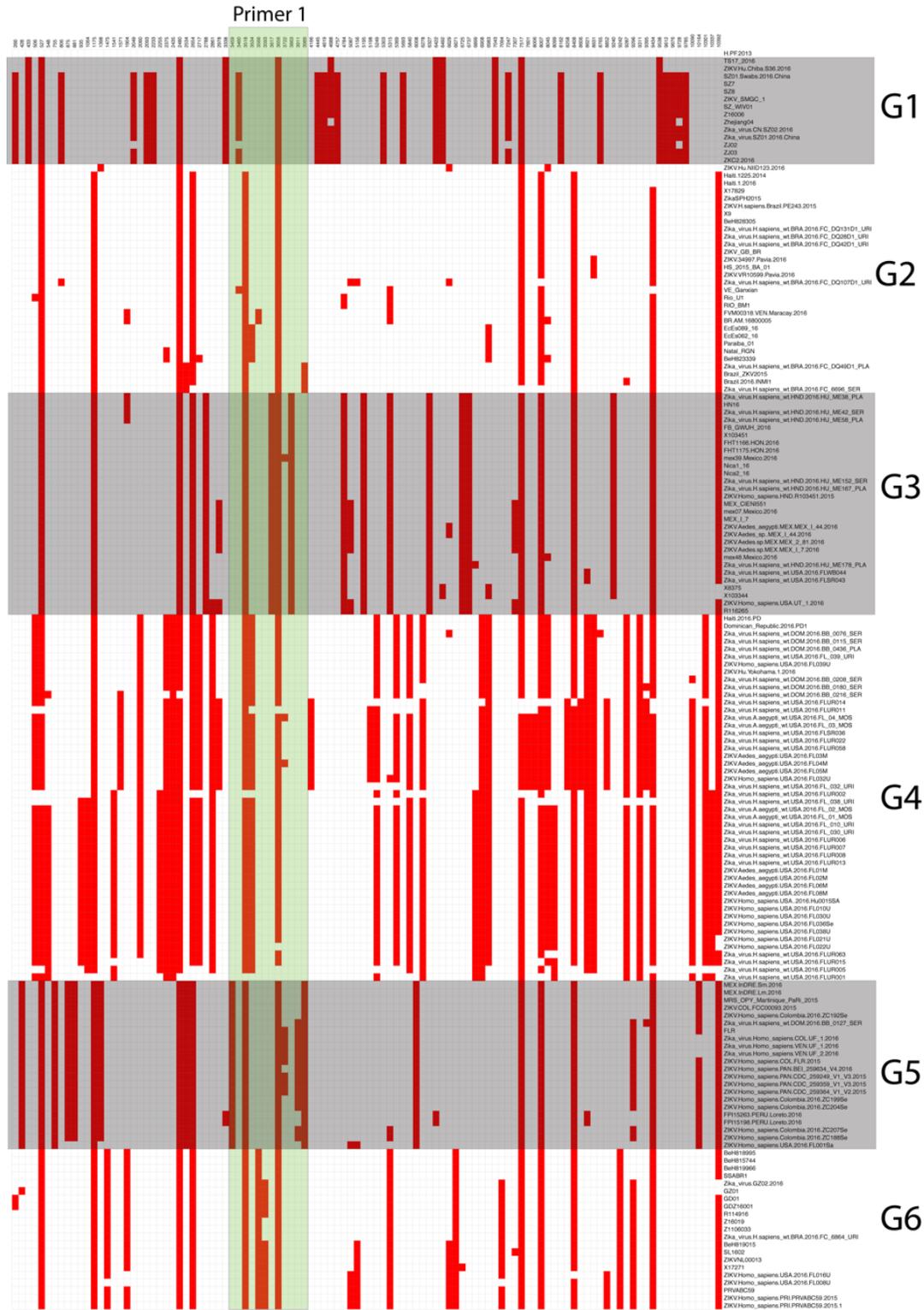
Figure 2.2.S6. A primer for the determination of viral groups. Primer 1 is selected and highlighted in region 3459-3989.

	G1	G2	G3	G4	G5	G6
G1	99.72±0.35	99.46±0.17	99.43±0.14	99.35±0.14	99.45±0.13	99.42±0.14
G2	99.46±0.17	99.65±0.14	99.61±0.11	99.53±0.11	99.62±0.11	99.60±0.10
G3	99.43±0.14	99.61±0.11	99.78±0.07	99.49±0.06	99.58±0.05	99.58±0.07
G4	99.35±0.14	99.53±0.11	99.49±0.06	99.83±0.08	99.50±0.04	99.47±0.06
G5	99.45±0.13	99.62±0.11	99.58±0.05	99.50±0.04	99.81±0.07	99.58±0.06
G6	99.42±0.14	99.60±0.10	99.58±0.07	99.47±0.06	99.58±0.06	99.73±0.09

Figure 2.2.S7. Genomic similarities (%) of ZIKV strains within and between the six groups.

2.3 POTENTIAL FOR TREATMENT AND A ZIKA VIRUS VACCINE

ABSTRACT

Zika virus (ZIKV) has only recently been exposed as a significant public health threat, and much of our limited knowledge of its pathogenesis and triggered immune responses were discovered in only the last few years. There are currently no ZIKV-specific therapeutics or vaccines available. This review seeks to bring the reader up-to-date with the latest developments in finding a way to combat this emerging infectious disease. Current strategies used for developing ZIKV vaccines or treatments follow proven methods used against other flaviviruses. Unfortunately, ZIKV carries many unique challenges, such as the need to target drugs and vaccines towards immunocompromised populations (pregnant mothers and fetuses), the risk of stimulating harmful immune responses (either autoimmune or antibody-dependent enhancement of infection in those with previous flavivirus exposure), frequently silent infection that may delay treatment and increase risk of transmission to others, and multiple routes of transmission (arthropod vector, sexual, bloodborne, and potentially other body fluids). Current medical recommendations are directed towards resolving symptoms and not the actual infection; however, ZIKV treatments and vaccines are in development. Vector control and travel restrictions to endemic areas may remain our only available interventions for quite some time.

2.3.a INTRODUCTION

Although Zika virus (ZIKV) was first discovered in 1947, it garnered little attention until this last decade due to an association with only mild symptoms and less than 20 reported human cases [1*]. However, at some point during its migration from Africa to Asia and subsequently the Pacific Islands, the virus mutated [2*, 3*, 4], changing the manifestations of its infection. Not only has it become more pathogenic, causing Guillain-Barré syndrome (GBS) and congenital malformations, but it has acquired (or perhaps we are simply just now uncovering) alternate routes of transmission not seen in any other mosquito-borne virus [1*, 5**]. These concerning developments have stirred public frenzy towards finding a means of halting the spread of this pandemic, as well as treatments to preserve the health of unborn children at risk of neurologic devastation [6*]. This review will highlight the key advancements made over the last 18 months in creating a ZIKV vaccine and in finding therapeutic options that neutralize the virus or counter its damaging effects.

The Foundations for building a ZIKV Vaccine

Fear over the potential global transmission of ZIKV is reminiscent of the public response to the Ebola virus (EBOV) outbreak in West Africa in 2014. However, unlike the high levels of morbidity and mortality associated with hemorrhagic fever viruses such as EBOV, ZIKV infection is symptomatic in only twenty percent of non-congenital cases, few of which even require hospitalization [1*]. As a category A bioterrorism agent, EBOV had the benefit of funding and years of former research by the U.S. defense department, enabling the rapid development and production of a vaccine within a year of the outbreak [7*]. By contrast, there have been very few studies on ZIKV infection prior to its appearance in the Yap Islands of Micronesia in 2007, and in French Polynesia in 2013 when it first became associated with GBS cases [5**, 6*]. It was only after a link between ZIKV maternal infection and adverse congenital outcomes was suspected [5**] that research in this field exploded. President Obama called for increased funding towards

ZIKV research, particularly vaccine development, some of which was necessarily re-appropriated from money originally intended to combat EBOV [8].

While we are faced with the challenge of creating a vaccine for a virus whose pathogenesis and host immune response we are only beginning to understand, we have the benefit of previous advancements made towards the study of related flaviviruses [9*]. Members of this genus all have linear, positive-sense nonsegmented RNA genomes inside enveloped capsids. The viruses are predominantly transmitted by either ticks or mosquitos, with a propensity to cause either neurologic or hemorrhagic disease. As the route of transmission is by insect bite to the skin with initial propagation in local keratinocytes, fibroblasts, and immature dendritic cells [10**], many flavivirus vaccines are administered subcutaneously to target those same cells. The first discovered and approved for human use is the 17D strain of yellow fever virus, which accumulated multiple stable and attenuating mutations throughout its genome after serial passaging in vitro. Researchers have also used 17D as a backbone to create chimeric vaccine strains for other flaviviruses, including Japanese encephalitis virus (JEV), DENV, West Nile and St. Louis encephalitis viruses, although only the first two have finished clinical trials and been approved for use in certain countries [11*, 12*, 13]. These recombinant vaccines replace the genes encoding the envelope (E) and pre-membrane (PrM) proteins of 17D with the ones from their respective viruses (the DENV vaccine includes genes for the 4 most common serotypes) [14*]. E is the surface antigen thought to be the predominant target of neutralizing antibody responses to flaviviruses. Other recombinant live DENV vaccines still in clinical trials include one that instead uses an attenuated DENV2 serotype virus as the backbone for the other three serotypes [15], and one comprised of four attenuated DENV strains [16]. An advantage of live attenuated vaccines is that they can induce long-lasting protective immune responses, similar to infection with the original virus.   Such a strategy could be used for a ZIKV vaccine, although it is not yet known whether ZIKV infection results in life-long immunity, nor whether distinct serotypes will emerge [3*]. Unfortunately, live vaccines also require more rigorous safety testing, as they can

be dangerous to immunocompromised hosts, such as pregnant women and their fetuses (ironically, the population we are most interested in protecting from ZIKV). Despite the promising results of the current DENV vaccine [14*], it still took over two decades of testing to win approval for human use [14*, 16, 17].

Other approaches to flavivirus immunization have also been tested in humans, such as intramuscular injection with purified formalin-inactivated tick-borne encephalitis virus, JEV or DENV, or with recombinant subunits of the DENV E and nonstructural protein 1 (NS1) proteins [11*, 17-19]. The NS1 of flaviviruses contains highly conserved areas within a secreted hexamer form that is involved in immune evasion, and has been targeted in other vaccines [20*]. Although less immunogenic than live viruses, inactivated and subunit vaccines can be tailored to deliver precise amounts of antigen, without the risk of live viruses in a multi-subtype formulation interfering with one another (a potential explanation for why the DENV tetravalent attenuated vaccine is not as effective against certain serotypes) [19]. Simultaneous immunization against all serotypes is important in preventing hemorrhagic complications that can arise in natural DENV infection when memory immune responses to one serotype later aggravate the disease caused by a different serotype [18]. Although this phenomenon is not well understood, antibody-dependent enhancement (ADE) of disease is a potential mechanism [19]. As ZIKV and DENV are closely related, there is a possibility that ZIKV infection in a population previously exposed to DENV or its vaccine may result in exacerbated disease [1*, 21**]. This has actually been proposed as a possible explanation for the GBS seen in 42 ZIKV-infected patients in French Polynesia, when DENV1 and 3 subtypes were also circulating [5**, 22*, 23*]. It is possible that a future ZIKV vaccine will therefore require co-administration with DENV vaccines to prevent ADE complications, however further study on this association is first warranted. It is also possible that the existing immunity to related flaviviruses in endemic regions will impact the effectiveness of a ZIKV vaccine, potentially boosting immunity, but also perhaps interfering with immunization [1*].

2.3.b RESULTS

ZIKV Vaccines in Development

The WHO conducted an analysis in March 2016 of all publically declared commercial, government and academic-led projects directed at ZIKV interventions, including vaccines (**Table 1**) [24**]. The list comprises multiple strategies, including vaccines using purified inactivated virus, nucleic acids, protein subunits, VLPs, and live recombinant attenuated viruses. However, most were still in the preclinical stages of development at the time of that posting, and Phase 1 clinical studies were not expected to begin until the end of 2016 [25*]. One factor potentially complicating vaccine studies is the lack of good animal models of ZIKV infection. Mice and other rodents are often the first subjects used to test vaccine effectiveness, as their immune responses have been well characterized and closely resemble those of humans. However, mice were found not to not show overt signs of ZIKV infection unless they were deficient in genes of the interferon (IFN) signaling pathway [26**, 27**] or tyrosine kinases [28*], which could confound the interpretation of protection studies. One group was able to induce ZIKV-neutralizing antibodies in susceptible mice after immunization with inactivated ZIKV virus or a DNA plasmid encoding the ZIKV prM-E proteins [29**]. To show that E-specific antibody titers also represent key immunologic correlates of protection in non-human primates, they immunized rhesus monkeys with either the purified inactivated vaccine (PIV), the DNA plasmid, or a recombinant rhesus adenovirus serotype 52 virus expressing those prM-E [30**]. While the adenovirus vaccine induced the strongest response, all three protected monkeys from subsequent ZIKV challenge, and the authors expected to begin clinical trials with the PIV in late 2016. To date, no other groups have published their preliminary findings, however the prevailing opinion among experts is that ZIKV vaccines will take several years to pass safety screening before being available for general distribution [1*, 6*]. While live attenuated vaccines may be the most immunogenic, other platforms will likely prove safer for pregnant women, an important high-risk population.

Currently Recommended Medical Interventions

Similar to other flavivirus infections, there are currently no virus-specific therapeutic interventions against ZIKV. After a 3-14 day incubation period, the infection remains asymptomatic in 80 percent of patients, while the remainder have an array of symptoms that can include low grade fever, maculopapular rash, myalgia, arthralgia, and conjunctivitis [6*]. Antihistamines can be used for pruritis, and fever and pain can be alleviated by acetaminophen (aspirin is not recommended in children due to risk of Reye's syndrome, and other nonsteroidal anti-inflammatory drugs can provoke hemorrhagic complications in cases of misdiagnosed DENV infection) [31*]. Symptoms are self-limited and usually resolve in 2-7 days, however a small subgroup of patients can progress to more serious complications, such as GBS. These would require hospitalization for monitoring and possible mechanical ventilation, intravenous immunoglobulin, and electrophoresis [23*]. Pregnant mothers with suspected ZIKV infection are recommended to undergo ZIKV rRT-PCR testing if the symptoms or exposure occurred within the previous 2 weeks (during which ZIKV is detectable in the blood, potentially up to 10 weeks)[32]. IgM testing is to be done first if the infection occurred after this period (IgG may give false positives due to cross-reactivity with other flaviviruses), but results should still be confirmed with PCR, DENV IgM, and/or PRNT [33**]. If there is a possibility of infection, serial fetal ultrasounds are recommended to detect possible congenital malformations, including microcephaly, cerebral calcifications, and brain atrophy [6*]. Children born with suspected congenital ZIKV syndrome are advised to be tested for ZIKV infection, as well as ophthalmologic exams, hearing screens, and periodic neurodevelopmental assessments [34**].

Preventative measures involve limiting exposure to mosquito vectors through restricted travel to endemic areas, protective clothing, insect repellants, and staying within air-conditioned environments. Blood banks have already begun screening for ZIKV contamination. As the virus has been sexually transmitted from asymptomatic hosts, can be detected in semen up to 6 months

after infection, and can replicate within the female vaginal tract, correct and consistent condom use is strongly recommended during oral, vaginal or rectal intercourse with anyone who is living in or has recently traveled to a ZIKV endemic region [1*, 35*].

Potential future ZIKV therapeutics

Although there are currently no treatments available against ZIKV, an improved understanding of its pathophysiology points us to potential drug targets. The viral envelope binds to multiple surface receptors on several host cell types to mediate entry, after which it replicates and induces autophagy to enhance its propagation [10**]. Type I interferons (IFN-I) can block viral replication and cell autophagy [10**, 36*], and mice deficient in the IFN receptor (IFNAR) have been found to be particularly sensitive to ZIKV infection [26**, 27**]. Interferon-induced proteins such as IFITM-1 and -3 have been shown to inhibit ZIKV replication and modulate its cytotoxicity [37*]. In addition, broad-spectrum antivirals such as ranpirnase and ribavirin, and antimalarials with antiviral properties such as chloroquine and amodiaquine, have been tested against flaviviruses and shown limited effectiveness [24**, 31*]. Most of these drugs are listed by the WHO as currently under investigation for ZIKV therapy (Table 2). In addition, 3-Methyladenine (3-MA), an inhibitor of autophagosome formation, was shown in vitro to reduce ZIKV copy numbers in infected fibroblasts [10**]. EGCG, a polyphenol found in green tea that has broad antiviral properties, was also found to inhibit ZIKV entry into cultured cells [38]. While these studies are preliminary, they demonstrate that progress is being made towards a cure, which could potentially be attainable long before a ZIKV vaccine.

Passive immunization could be considered for ZIKV treatment or short-term protection during high-risk periods of exposure [24**, 31*]. Polyclonal serum from immune donors was used in EBOV-infected patients during that recent outbreak. Although its effectiveness was unclear, given the present lack of alternatives, convalescent serum could represent a relatively safe option for use during pregnancy to accelerate ZIKV clearance and potentially reduce vertical transmission

64

[24**]. Virus-specific monoclonal antibodies (mAbs) are an improvement to this approach, and have shown promise in animal models against other flaviviruses. Progress has been made in testing mAbs from ZIKV-infected donors and screening them for cross-reactivity with DENV, and a promising candidate has already shown protection in mice [39**]. While mAb development uses a more straightforward strategy than vaccines, given ZIKV's propensity to induce adverse antibody reactions such as ADE and GBS, rigorous testing will still be necessary [23*].

Therapies specifically aimed at stopping the teratogenic effects of ZIKV are difficult to study, given that the mechanisms by which the virus crosses the placenta and causes fetal neuronal tissue destruction have only just been elucidated. Experiments in mice have shown that ZIKV infects cortical progenitor cells, inducing apoptosis and autophagy that leads to microcephaly [28**]. A screen of ~6000 compounds for ZIKV inhibition in infected neuronal cells identified emricasan, niclosamide, and PHA-690509 as promising products [40*]. A study examining how ZIKV interacts with host receptor TIm1 for cell entry found that duramycin can block this process, resulting in reduced infection of placental cells in vitro [41*]. Although mouse models of ZIKV-induced microcephaly exist for which these drugs could be tested, screening for safety and effectiveness in pregnant women will likely take years.

			Clinical Trial Phase I			
Company	Base	Technologies	Start date	End date*	Status and timelines**	Collaboration
Bharat	India	Inactivated; VLP with prM/E	2017[+]	2018	Preclinical work completed	
Bio-Manguinhos/Fiocruz	Brazil	Inactivated; YF17D chimeric; VLP; DNA			Preclinical work initiated	Under consideration
Butantan	Brazil	Live dengue recombinant; Inactivated			Preclinical work initiated	US NIH
GeoVax	U.S.	MVA expressing VLP; VLP			Preclinical work initiated	University of Georgia
GlaxoSmithKline	U.K	Self-amplifying mRNA (SAM); Inactivated			Preclinical work initiated	US NIH
Harvard University	U.S.	VSV vectored			Preclinical work initiated	
Hawaii Biotech	U.S.	Recombinant proteins			Preclinical work initiated	
Jenner Institute	U.K	Chimpanzee adenovirus vectored			Preclinical work initiated	
InOvio/GeneOne	U.S./Korea	DNA (GLS-5700)	Jul 2016	Nov 2016	Phase 1: active	Université Laval; UPenn
Institut Pasteur	France	Lentivirus vectored; measles vectored			Preclinical work initiated	Themis
Moderna	U.S	mRNA vaccine			Preclinical work initiated	US DHHS; US BARDA
NewLink	U.S.	Inactivated			Preclinical work initiated	Merck
Novavax	U.S.	E protein nanoparticles			Preclinical work initiated	
PaxVax	U.S	Inactivated			Preclinical work initiated	
Protein Sciences	U.S	Recombinant E protein			Preclinical work initiated	
Replikins	U.S.	Synthetic replilink peptides			Preclinical work initiated	
Sanofi Pasteur	France	Live attenuated (ChimeriVax)	2016[+]	2017	Preclinical work completed	WRAIR
Sanofi Pasteur	France	Inactiviated	2018[+]	2019	Preclinical work initiated	WRAIR; US BARDA
SLU vaccine center	U.S.	Inactivated (ZPIV)	Nov 2016	Nov 2017	Preclinical work completed	
Takeda	Japan	Inactivated			Preclinical work initiated	US BARDA
Themis Bioscience	Austria	Measles vectored			Preclinical work initiated	Institut Pasteur
US CDC	U.S.	DNA plasmid expressing VLP; live recombinant adenovirus			Preclinical work initiated	
US NIAID	U.S.	DNA (VRC-ZKADNA085-00-VP)	Jul 2016	Dec 2017	Phase 1: recuiting	Various
US NIAID	U.S.	Live attenuated; Live recombinant VSV			Preclinical work initiated	Various
UTMB/Instituto Butantan	U.S./Brazil	Live attenuated			Preclinical work initiated	
Valneva	France	Inactivated			Preclinical work initiated	

Figure 2.3.1. ZIKV Vaccines in Development. In addition, the following institutions have communicated about their active consideration of the field or have committed planning/discovery stage activities: CureVac, Johnson & Johnson, Oxford University, Pfizer, Profectus Biosciences, Sementis, Sinergium. [+] Estimated start date of clinical trials, * Estimated Primary Completion Date, ** Preclinical work refers to animal studies, Table modified from [24].

Therapeutic	In vivo activity against ZIKV or other flaviviruses	In vitro data (Effective concentration, specificity index, cell type)	Safety/ Use in pregnant women	Availability/ Feasibility
Amodiaquine	unknown	DENV EC90=2.7uM, SI ~10 in BHK-21 cells	unlikely teratogenic; occasional agranulocytosis, neutropenia, hepatotoxicity SAE in long-term use in some CYP2C8 gene variants	common antimalarial; possible benefit against EBOV
Chloroquine	DENV: no decrease in adult viremia	DENV: 0.5 ug/ml in Vero cells. No effect in C6/36 cells	Safe	common antimalarial
Ribavirin	DENV: not effective in NHPs; YF: increased hamster survival	ZIKV: EC50 = 140 ug/ml Vero, SI >55; DENV: EC50 = 20 ug/ml Vero, SI >400; YF: EC50=42ug/ml Vero, SI =174	Teratogenic	readily available; broad use antiviral
Interferon a	JEV: no effect in infants	JEV: EC50=4.8 IU/ml Vero; ZIKV: EC50= 34 IU/ml Vero		efficacy against hepatitis viruses
BCX4430 (Biocryst, USA)	YF: increased hamster survival	YF: EC50= 8.3 ug/ml Vero SI = ~5; DENV: EC50=13ug/ml; WNV: EC50=16 ug/ml	Phase I safety completed, no information on teratogenicity	
GS-5734 (Gilead, USA)	EBOV: reduced mortality and pathology in infected NHPs	EBOV: inhibits viral replication in multiple human cell types; ZIKV: under investigation	Phase I safety completed, no information on teratogenicity	
NITD008	DENV2: decreased viremia/mortality in mice	DENV: EC50 = 3uM; WNV: EC50=5 uM; YF: EC50=3uM	no human safety data	likely long time to human safety data
Monoclonal antibodies	ZIKV: complete protection in mice		likely safe but possible antibody-mediated pathology	cost and other considerations limit widespread use
Emricasan		Protected cultured human neural cells from ZIKV-induced cell death	Phase II for liver protection from HCV showed no SAE; teratogenicity unknown	being studied for protection from hepatitis C virus (HCV) liver pathology
Niclosamide		Inhibited JEV replication in BHK21 cells; inhibited ZIKV replication in human neural cells	Pregnancy category B	common teniacide
PHA-690509		Inhibited ZIKV replication in cultured human neural cells	Teratogenicity unknown	was in Phase I trials for cancer therapy
Duramycin		DENV, WNV: reduced infection in A549 and Vero; ZIKV: reduced infection in multiple placental cell types	hemolytic at high concentrations; teratogenicity unknown	aerosolized form in clinical trials for cystic fibrosis
3-Methyladenine (3-MA)		reduced ZIKV viral copies in infected skin fibroblast cell line	Teratogenicity unknown	authophagy inhibitor; used in chemotherapy studies
EGCG		inhibited ZIKV infection of Vero cells	crosses placenta but no teratogenic effects seen in rats	controversial antiviral/dietary supplement; poor bioavailability
Ranpirnase	EBOV: decreased viremia/mortality in mice	inhibited infections of Vero cells with EBOV and undisclosed cells with ZIKV	Was in phase III for mesothelioma; no information on teratogenicity	induces apoptosis in malignant cells; inhibits papillomaviruses

Figure 2.3.2. Potential ZIKV Therapeutics. Table modified from [24], with additional entries from [10, 31, 38, 39, 40, 41] abbreviations not used in text: SAE = serious adverse event; NHP = nonhuman primate.

2.3.c DISCUSSION

We must rely on lessons learned from related flaviviruses to quickly design new therapeutic and prophylactic ZIKV interventions. Three vaccines have already shown promise in monkeys, and many more are in development. ZIKV antivirals are still in early testing, although monoclonal antibodies have worked for other infections and could be easier to screen and produce. Each endeavor has the added challenge of proving safety in pregnant women, and accounting for possible ZIKV induction of adverse antibody responses leading to GBS and potentially ADE in populations with preexisting flavivirus immunity. Therefore, immunizations and treatments against ZIKV will require prolonged testing.

- Zika virus (ZIKV) is an emerging global public health threat, with no treatments or vaccines currently available.
- Our understanding of ZIKV is founded on limited research from only the last few years, therefore medical intervention strategies currently draw on knowledge of related flavivirus infections.
- The developing fetus is most at risk of ZIKV complications, therefore pregnant women must be targeted for treatments and vaccines; however this presents challenges for ensuring both efficacy and safety in this immunocompromised population.
- Gaps in our ZIKV knowledge base, animal model limitations, and concern for adverse immune responses and teratogenic effects, all pose challenges to drug developers, and could prolong the wait before a ZIKV vaccine or treatment becomes available.

REFERENCES AND RECOMMENDED READING

Papers of particular interest, published within the annual period of review, have been highlighted as:

* of special interest

** of outstanding interest

REFERENCES

1.   * Lazear HM, Diamond MS. Zika Virus: New Clinical Syndromes and Its Emergence in the Western Hemisphere. J Virol. 2016;90(10):4864-75.

This is a comprehensive review of ZIKV's history and new developments. The authors, experts in arboviruses and immunology, also discuss issues in ZIKV vaccine development and adverse antibody responses.

2.   * Wang L, Valderramos SG, Wu A, et al. From Mosquitos to Humans: Genetic Evolution of Zika Virus. Cell Host Microbe. 2016;19(5):561-5.

This article analyzes the genetic changes ZIKV has accumulated over time and their potential effects on virulence and fitness.

3.   * Ye Q, Liu ZY, Han JF, Jiang T, et al. Genomic characterization and phylogenetic analysis of Zika virus circulating in the Americas. Infect Genet Evol. 2016;43:43-9.

This study compares the ZIKV genome with other arboviruses and notes regions of conservation and immunogenicity important for vaccine development.

4.   Faye O, Freire CC, Iamarino A, et al. Molecular evolution of Zika virus during its emergence in the 20(th) century. PLoS Negl Trop Dis. 2014;8(1):e2636.

5.	** World Health Organization. Zika Situation Report Neurological Syndrome and Congenital Anomalies. World Health Organization, 2016. 05 February Report.

This was the initial announcement by the WHO acknowleding the association between ZIKV infection and congenital microcephaly. This significantly raised international concern and spurred new ZIKV funding and research.

6.	* Focosi D, Maggi F, Pistello M. Zika Virus: Implications for Public Health. Clin Infect Dis. 2016;63(2):227-33.

This review discusses ZIKV transmission risks and current measures for prevention.

7.	* Hoyt K, Hatchett R. Preparing for the next Zika. Nat Biotechnol. 2016;34(4):384-6.

This commentary discusses the obstacles of rapid vaccine development in response to recent pandemics.

8.	Gostin LO, Hodge JG, Jr. Is the United States Prepared for a Major Zika Virus Outbreak? JAMA. 2016;315(22):2395-6.

9.	* Tripp RA, Ross TM. Development of a Zika vaccine. Expert Rev Vaccines. 2016;15(9):1083-5.

This editorial discusses previous successes in flavivirus immunizations, proposing those same avenues for ZIKV.

10.	** Hamel R, Dejarnac O, Wichit S, et al. Biology of Zika Virus Infection in Human Skin Cells. J Virol. 2015;89(17):8880-96.

This paper reveals breakthrough findings in the mechanisms of ZIKV cellular infection. It also tests the effectiveness of antivirals and autophagy inhibitors on ZIKV infection in cultured cells.

11.   * Weaver SC, Costa F, Garcia-Blanco MA, et al. Zika virus: History, emergence, biology, and prospects for control. Antiviral Res. 2016;130:69-80.

This lengthy review summarizes the gaps in our preparedness against ZIKV and points out issues of concern for drug and vaccine development.

12.   * Leibovitch EC, Jacobson S. Vaccinations for Neuroinfectious Disease: A Global Health Priority. Neurotherapeutics. 2016;13(3):562-70.

This article discusses the successes and challenges of vaccination against a large variety of different neurological infections,.

13.   Monath TP, McCarthy K, Bedford P, et al. Clinical proof of principle for ChimeriVax: recombinant live, attenuated vaccines against flavivirus infections. Vaccine. 2002;20(7-8):1004-18.

14.   * Wilder-Smith A, Yoon IK. Edging closer towards the goal of a dengue vaccine. Expert Rev Vaccines. 2016;15(4):433-5.

This editorial objectively outlines the successes and limitiations of the currently approved live attenuated chimeric DENV vaccine. This vaccine is a model being pursed for ZIKV, to which DENV is closely related.

15.   Osorio JE, Velez ID, Thomson C, et al. Safety and immunogenicity of a recombinant live attenuated tetravalent dengue vaccine (DENVax) in flavivirus-naive healthy adults in Colombia: a randomised, placebo-controlled, phase 1 study. Lancet Infect Dis. 2014;14(9):830-8.

16.   Durbin AP, Kirkpatrick BD, Pierce KK, et al. A 12-Month-Interval Dosing Study in Adults Indicates That a Single Dose of the National Institute of Allergy and Infectious Diseases Tetravalent Dengue Vaccine Induces a Robust Neutralizing Antibody Response. J Infect Dis. 2016;214(6):832-5.

17.   Schwartz LM, Halloran ME, Durbin AP, Longini IM, Jr. The dengue vaccine pipeline: Implications for the future of dengue control. Vaccine. 2015;33(29):3293-8.

18.   Coller BA, Clements DE, Bett AJ, et al. The development of recombinant subunit envelope-based vaccines to protect against dengue virus induced disease. Vaccine. 2011;29(42):7267-75.

19.   Fernandez S, Thomas SJ, De La Barrera R, et al. An adjuvanted, tetravalent dengue virus purified inactivated vaccine candidate induces long-lasting and protective antibody responses against dengue challenge in rhesus macaques. Am J Trop Med Hyg. 2015;92(4):698-708.

20.   * Rastogi M, Sharma N, Singh SK. Flavivirus NS1: a multifaceted enigmatic viral protein. Virol J. 2016;13:131.

This review describes the roles of NS1 in flavivirus cell invasion and immune evasion, and its importance as a vaccine target.

21.   ** Dejnirattisai W, Supasa P, Wongwiwat W, et al. Dengue virus sero-cross-reactivity drives antibody-dependent enhancement of infection with zika virus. Nat Immunol. 2016;17(9):1102-8.

This study supports the feared possibility that DENV and ZIKV crossreactive antibodies can lead to enhanced disease. Exposure of ZIKV to human sera with antibodies to DENV was able to augment the infection of cultured cells.

22.   * Garcia E. YE, Nishino E., Millot V., et al. Zika virus infection: global update on epidemiology and potentially associated clinical manifestations. Wkly Epidemiol Rec. 2016;91(7):73-81.

This report by the WHO describes the rise of GBS seen in ZIKV cases around the world.

23.   * Malone RW, Homan J, Callahan MV, et al. Zika Virus: Medical Countermeasure Development Challenges. PLoS Negl Trop Dis. 2016;10(3):e0004530.

This review descibes interventions in development against ZIKV and related flaviviruses, and their limitations.

24.   ** World Health Organization. Current Zika Product Pipeline. World Health Organization, 2016. 03 March Report.

This report by the WHO lists the status of vaccines and drugs in development from multiple projects worldwide. The list was only current as of March 2016, but we have updated it in tables 1 and 2 of this review.

25.   * Fauci AS, Morens DM. Zika Virus in the Americas--Yet Another Arbovirus Threat. N Engl J Med. 2016;374(7):601-4.

This review characterizes the ZIKV pandemic compared to those of other arboviruses and discusses the urgency of finding a vaccine.

26.   ** Lazear Helen M, Govero J, Smith Amber M, et al. A Mouse Model of Zika Virus Pathogenesis. Cell Host & Microbe. 2016.

This study was the second to describe an animal model for ZIKV infection with contemporary viruses. It used several mutant strains of mice to characterize the importance of IFN signaling.

27. ** Aliota MT, Caine EA, Walker EC, et al. Correction: Characterization of Lethal Zika Virus Infection in AG129 Mice. PLoS Negl Trop Dis. 2016;10(5):e0004750.

This breakthrough study was the first to describe an animal model for studying infection with contemporary ZIKV strains. It also characterized the pathology seen in these mice.

28. ** Cugola FR, Fernandes IR, Russo FB, et al. The Brazilian Zika virus strain causes birth defects in experimental models. Nature. 2016;534(7606):267-71.

This study presented a mouse model for congenital ZIKV infection, and described the pathology seen. In addition, they characterized the ZIKV infection of human brain organoids.

29. ** Larocca RA, Abbink P, Peron JP, et al. Vaccine protection against Zika virus from Brazil. Nature. 2016;536(7617):474-8.

This is the first publication that details the findings of a ZIKV vaccine in mice. Both a PIV and DNA plasmid vaccine were tested and individually conferred full protection to mice, which was noted to be antibody and not T-cell mediated.

30. ** Abbink P, Larocca RA, De La Barrera RA, et al. Protective efficacy of multiple vaccine platforms against Zika virus challenge in rhesus monkeys. Science. 2016.

This article describes the subsequent immunization of monkeys with the vaccines from [29], as well as an adenovirus-delivered live ZIKV vaccine. All three were individually able to confer full protection in this model, a promising step towards proving efficacy in humans.

31. * Fellner C. Zika Virus: Anatomy of a Global Health Crisis. P T. 2016;41(4):242-53.

This review focuses on the current interventions against ZIKV and the ones in development.

32.   Driggers RW, Ho CY, Korhonen EM, et al. Zika Virus Infection with Prolonged Maternal Viremia and Fetal Brain Abnormalities. N Engl J Med. 2016;374(22):2142-51.

33.   ** Oduyebo T, Igbinosa I, Petersen EE, et al. Update: Interim Guidance for Health Care Providers Caring for Pregnant Women with Possible Zika Virus Exposure - United States, July 2016. MMWR Morb Mortal Wkly Rep. 2016;65(29):739-44.

This report describes the updated CDC guidelines outlining the screening and care of pregnant women at risk of ZIKV infection and transmission to the fetus.

34.   ** Russell K, Oliver SE, Lewis L, et al. Update: Interim Guidance for the Evaluation and Management of Infants with Possible Congenital Zika Virus Infection - United States, August 2016. MMWR Morb Mortal Wkly Rep. 2016;65(33):870-8.

This report by the CDC describes the recommended screening and management of children with suspected congenital ZIKV infection.

35.   * Brooks JT, Friedman A, Kachur RE, et al. Update: Interim Guidance for Prevention of Sexual Transmission of Zika Virus - United States, July 2016. MMWR Morb Mortal Wkly Rep. 2016;65(29):745-7.

This report describes the known cases, risks, and necessary preventative practices for ZIKV sexual transmission.

36.   * Grant A, Ponia SS, Tripathi S, et al. Zika Virus Targets Human STAT2 to Inhibit Type I Interferon Signaling. Cell Host Microbe. 2016;19(6):882-90.

This article describes how ZIKV NS5 interacts with human STAT2 to interfere with IFN signaling. The fact that it doesn't interact with mouse STAT2 could explain why immune competent mice do not experience ZIKV disease.

37.   * Savidis G, Perreira JM, Portmann JM, et al. The IFITMs Inhibit Zika Virus Replication. Cell Rep. 2016;15(11):2323-30.

The mutation studies in this article demonstrate the importance of specific interferon-stimulated genes in preventing ZIKV infection, and suggests that enhacing their effects could be a means of protection against ZIKV.

38.   Carneiro BM, Batista MN, Braga AC, et al. The green tea molecule EGCG inhibits Zika virus entry. Virology. 2016;496:215-8.

39.   ** Stettler K, Beltramello M, Espinosa DA, et al. Specificity, cross-reactivity, and function of antibodies elicited by Zika virus infection. Science. 2016;353(6301):823-6.

This study characterizes the epitopes that generate cross-reactive antibodies between DENV and ZIKV. One particular antibody was also shown to protect mice against ZIKV infection.

40.   * Xu M, Lee EM, Wen Z, et al. Identification of small-molecule inhibitors of Zika virus infection and induced neural cell death via a drug repurposing screen. Nat Med. 2016.

This article indentifies promising ZIKV drugs from a screen of over 6000 compounds for antiviral or neuroprotective properties. These drugs showed efficacy in ZIKV-infected human neural cells.

41.   * Tabata T, Petitt M, Puerta-Guardo H, et al. Zika Virus Targets Different Primary Human Placental Cells, Suggesting Two Routes for Vertical Transmission. Cell Host Microbe. 2016;20(2):155-66.

The results of this study suggest that ZIKV can infect the fetus through both a placental and a paraplacental route, based on the cell types permissive to infection *in vitro*. Duramycin was also shown to prevent ZIKV infection of these cells.

CHAPTER 3

GENOME-WIDE MUTAGENESIS OF INFLUENZA VIRUS

3.1 FUNCTIONAL GENOMICS REVEALS LINKERS CRITICAL FOR INFLUENZA POLYMERASE

ABSTRACT

Influenza virus mRNA synthesis by the RNA-dependent RNA polymerase involves binding and cleavage of capped cellular mRNA by the PB2 and PA subunits, respectively, and extension of viral mRNA by PB1. However, the mechanism for such a dynamic process is unclear. Using high-throughput mutagenesis and sequencing analysis, we have not only generated a comprehensive functional map for the microdomains of individual subunits, but have also revealed the PA linker to be critical for polymerase activity. This PA linker binds to PB1 and also forms ionic interactions with the PA C-terminal channel. Nearly all mutants with five amino acid insertions in the linker were non-viable. Our model further suggests that the PA linker may play an important role in the conformational changes that occur between stages that favor capped mRNA binding and cleavage and those associated with viral mRNA synthesis.

3.1.a INTRODUCTION

Influenza A virus is a major public health problem, infecting as many as 500 million people a year world-wide and causing more than 500,000 deaths (1). RNA-dependent RNA polymerase (RdRp) consists of the PA (polymerase acidic), PB1 (polymerase basic 1) and PB2 (polymerase basic 2) subunits, has been the focus of a great deal of research (2-6). The nearly complete crystal structure of the bat influenza polymerase was recently published, providing a physical map of the subunits (7). Although it is a promising drug target, its multiple functions and how they relate to specific structures within the heterotrimer have not yet been fully elucidated.

Once the virus has invaded the host cell nucleus, the RdRp transcribes the vRNA genome into capped and polyadenylated mRNAs. The cap-binding domain of the PB2 subunit binds the 5′ 7-methyl-guanosine cap structure of a cellular pre-mRNA (8, 9), and subsequently, the N-terminal region of the PA subunit endonuclease (PAn) is thought to cleave off and snatch the 10–15 nucleotides downstream of the cap (10, 11), allowing them to serve as a primer for mRNA synthesis on the vRNA template through a conformational change in PB2. The ends of the single-stranded vRNAs then re-associate to restore the dsRNA promoter structure, leading to PB1 initiating polyadenylation of the viral mRNA (8, 12). The N-terminus of PB2 (PB2n) is also responsible for binding with the C-terminal region of PB1 (PB1c) and the replication of the vRNA genome, using cRNA intermediates as templates (8, 13, 14). The C-terminus of PB2 (PB2c) carries a bipartite nuclear localization signal (NLS), which is thought to be responsible for α5 importin binding (15, 16). However, the functions of many microdomains of the of the influenza virus's heterotrimeric replication machinery remain unclear.

Our approach has involved using a transposon mutagenesis system to generate short in-frame insertions and create ~10^5 different insertion clones from a single reaction for each subunit gene. This non-biased high-throughput analysis provides a rapid and effective method to pin-point microdomains of individual subunits for further functional studies. Here we have shown the first structure-function analysis of the entire influenza polymerase complex, including catalytic regions and nine linkers connecting different microdomains. Information we acquired from this approach

has led us to discover the function of an extended PA linker, which binds to PB1 and also forms ionic interactions with the PA C-terminal (PAc) channel. Our data has also highlighted the critical and previously uncharacterized role influenza A virus linkers play in coordinating the spatial interactions and functions of the polymerase complex.

The elucidation of the structure-function relationship of the influenza replication machinery is indispensable not only for a profound mechanistic understanding of the virus's infectivity, activity and pathogenesis, but also for designing more effective therapeutics against this recurring pathogen that has caused and will continue to cause significant morbidity and mortality.

3.1.b RESULTS

Generations of high-throughput mutagenesis profiling of the PA, PB1 and PB2 proteins

We created three mutant libraries comprised of $>10^5$ random in-frame insertions of a 15 nucleotide sequence, 5'-NNNNNTGCGGCCGCA-3', in the PB1, PB2 and PA gene segments of influenza A/WSN/1933 H1N1 (WSN) using a Mu-transposon mutagenesis method previously described (17). We used the reverse genetics system of influenza to transfect mutant plasmid libraries of each polymerase gene segment with the complementary 7 segments of wild-type (wt) WSN into HEK293T cells, creating a mutant virus library (Figs. 1A and S2). High-throughput genotyping of each library shows a low level of insertion tolerance, which is expected given the vital role of polymerase genes. However, we found that ~50% of the clones that were observed in the passage 1 population remained in passage 3 (Figs. 1B-D and Table S5; peaks represent mean fluorescence intensity [MFI]). One advantage of using Mu-transposon mutagenesis is that it generates random insertions of 5 amino acids in all 3 reading frames without producing stop codons. To verify this, we selected 152 single insertion mutant clones from all 3 libraries and examined polymerase activity (Table S2-4). We found that the viability and the polymerase activity of the single clones strongly supported the high-density mutagenesis results (Fig. S3 and 4), and 40/42 single clones could express protein at a wt level (data not shown). This data provides the fundamental justification that our insertional mutation method can be used as a high-throughput loss-of-function analysis tool.

After three rounds of passaging in MDCK cells, we mapped the viable clones of the PA, PB1 and PB2 libraries onto a linear schematic and the published 3D protein structure (Figs. 2A-C) (7, 18). The regions we identified where viable insertion mutants could not be recovered matched the previously identified functional domains, confirming their importance. For Instance, we confirmed that the PB1 library recovered no viable mutations in the vRNA binding regions (233-249 and 560-574) or the conserved motifs preA to D (225-252,300-317,403-426,440-452 and 475-487) (19-21). Several mutants in the PB1c-PB2n binding region were recovered from passage 1, but only one remained by passage 3. The newly discovered PB1 priming site and the

β hairpin loop (7), which were suggested to interact with the template and vRNA, tolerated only one insertion mutant. However, we found many viable mutations outside of those functional regions (Fig. S4B). Several mutants in the single-partite nuclear localizing signal (NLS) regions at amino acid positions 205, 206, 207 and 210 survived through all passages.

In the PB2 schematic shown in figure 2C, the PB1c- and NP-binding sites are both located at PB2n and overlap each other. No viable mutants were discovered in the PB2n-PB1c binding domain. The newly discovered lid domain, which may open to allow template exiting, tolerates few mutations at the non-flexible α helixes (7). No mutants were recovered within the two suggested cap-binding regions. Mutations in close proximity to amino acid 627, which plays a crucial role in viral transmission, also failed to produce virus. In contrast to the PB1 NLS region, no viable mutants were found in the bipartite NLS regions in PB2c (736-739 and 752-755, respectively).

In the PA schematic shown in figure 2A, PAn, which is responsible for endonuclease activity, can surprisingly tolerate multiple insertions, but not at positions K102, D108 or K134. What is known is that the function of the endonuclease relies on a negatively charged sequence in these residues (11, 22). The high tolerance we observed also correlates with Hiromoto *et al's* finding that PAn has been the most mutated region in the PA subunit over the course of evolution (23). In contrast to the many viable mutants in the endonuclease domain, we found no mutants in the 'mouth' region, particularly the α11, α21 and α22 helixes, or the 'brain' of PAc (291-716), which consists of the channel domain and the RNA-binding groove (24). These findings demonstrate that genome-wide insertion mutagenesis can provide important clues in identifying functionally important regions in less characterized proteins.
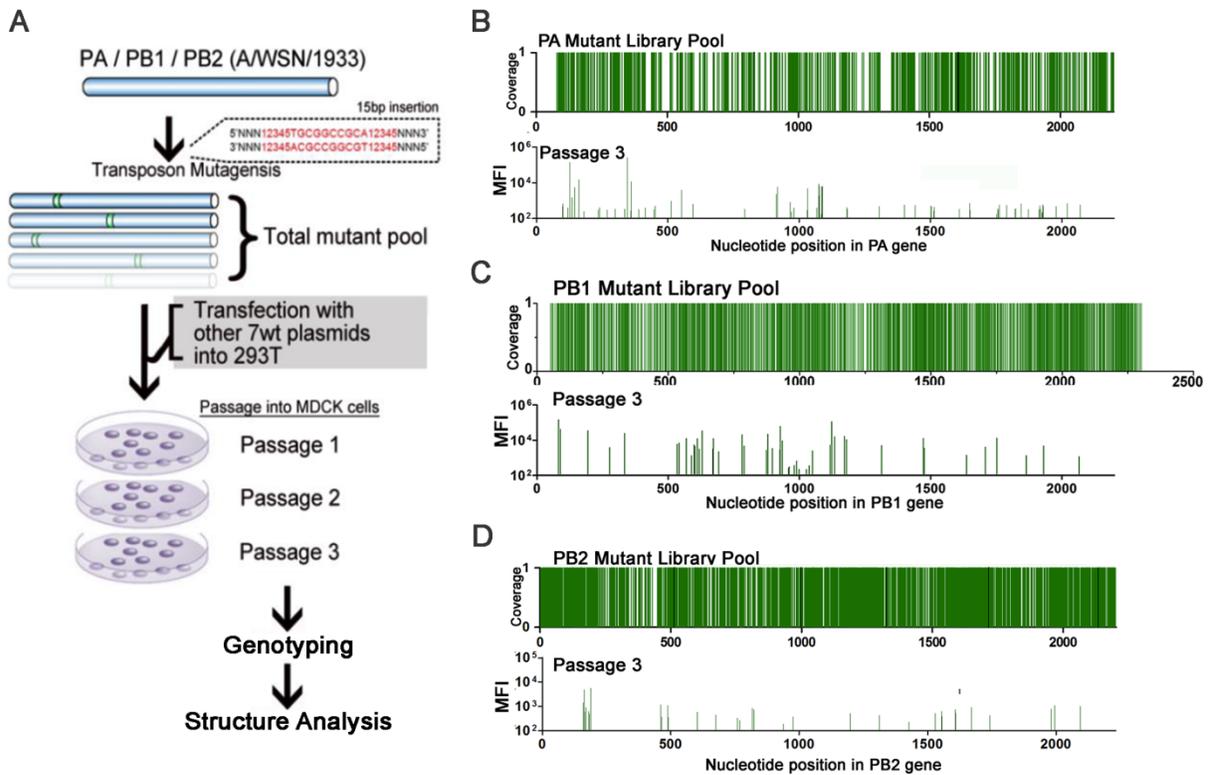
Figure 3.1.1. Generations of high-throughput mutagenesis profiling of the PA, PB1 and PB2 proteins. (A) The flowchart for the generation of a PB1, PA and PB2 mutant library pool. Peaks from the PA **(B)**, PB1 **(C)**, and PB2 **(D)** plasmid library pool demonstrate the total coverage of each subunit, which is co-transfected with 7 wt WSN plasmids in 293T cells and 3 passages in MDCK cells. The *in vitro* replication efficiency of the entire mutant pool was determined by genotyping of the cell lysates, as mean fluorescence intensity [MFI]. The list of the all viable mutants of each mutant library from passage 1-3 can be found in Table S5.
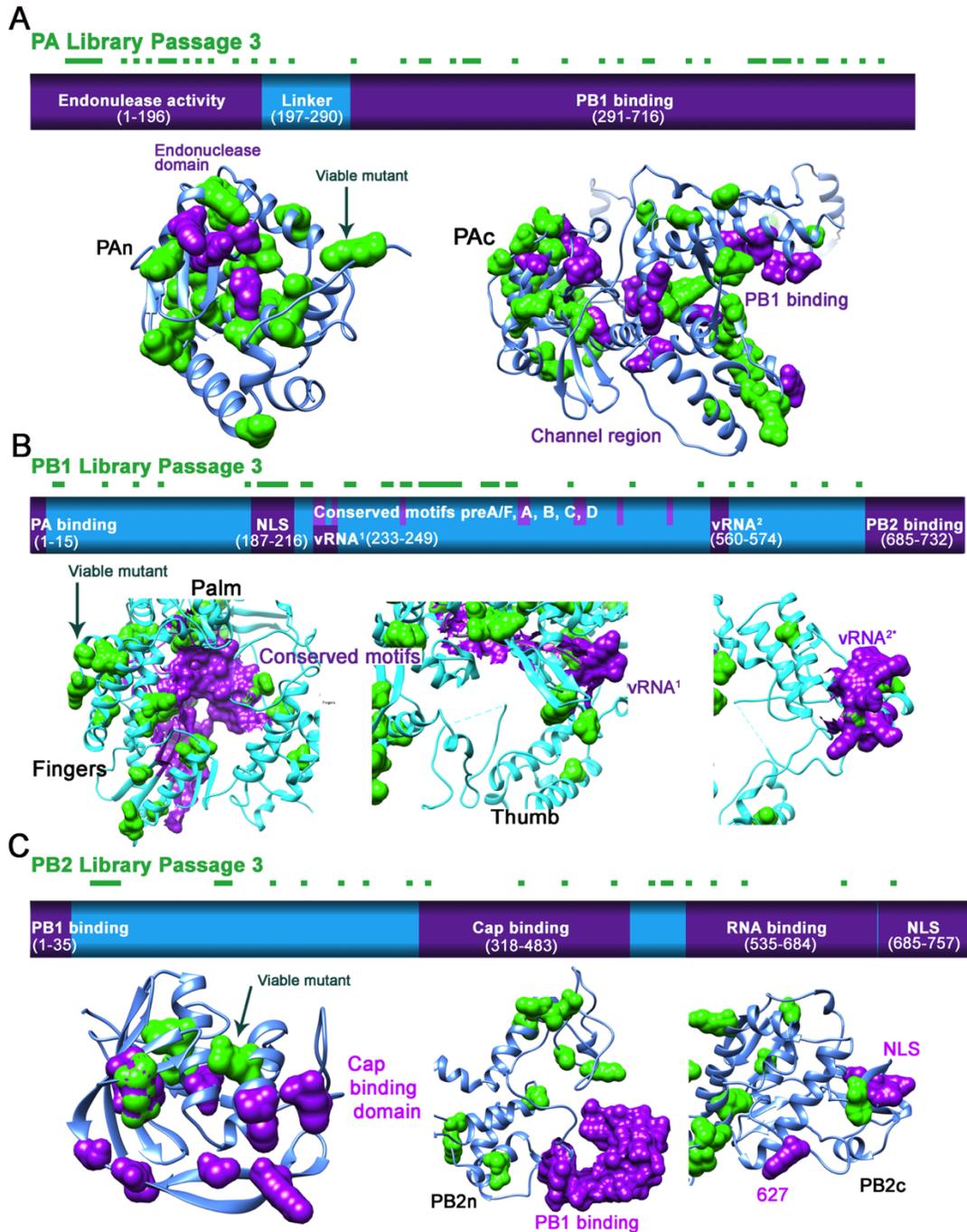
Figure 3.1.2. Structure and functional characterization of the influenza polymerase complex. Graphical illustration of PA **(A)**, PB1 **(B)** and PB2 **(C)** functional domains (based on previous literature shown in purple) overlapped with passage 3 mutant library as square dots on top of the bar (green). Using the same color scheme, we mapped all of viable and non-viable (light blue) mutants from passage 3 into the recently identified influenza crystal structure (PDB ID: 4WSB) (7, 18). The viable and functional regions were highlighted using the surface display.

Characterization of critical PB1 microdomains using insertion mutagenesis

The template binding and catalytic channel (TBC) is crucial for binding template, substrates and cofactors, as well as catalyzing the nucleotidyl transferase reactions (21, 25-27). The third passage of the PB1 library revealed that residues inside the TBC could not tolerate insertions, in contrast to residues on the surface away from the TBC (data not shown). We hypothesized that the TBC is involved in RNA synthesis. Therefore, we isolated five individual mutant clones from the PB1 mutant plasmid library (Table S2) with insertions in the TBC and tested their RdRp activity and mRNA, vRNA and cRNA synthesis efficiency using a mini-replicon assay (Fig. 3A). We found that these mutant clones lacked any RdRp activity, were unable to proceed with any step in RNA synthesis, and were unable to produce viable virions when transfected, supporting the critical role of this channel.

In addition, our mutational analysis uncovered two interesting non-viable PB1 mutants with insertions at amino acid positions 756 and 766, which are at the end and outside of the coding region, respectively (Fig. 3B). When we isolated these mutants and tested their RdRp activity, we found levels similar to wt PB1, suggesting that polymerase functions related to heterotrimeric interactions and RNA binding were unaffected. However, results from a viral packaging assay showed that these mutants had defects in viral assembly (Figs. 3C and 3D). This additional involvement in later stages of viral replication suggests that the PB1 3' UTR may play a more crucial role than previously thought in viral packaging (28, 29).
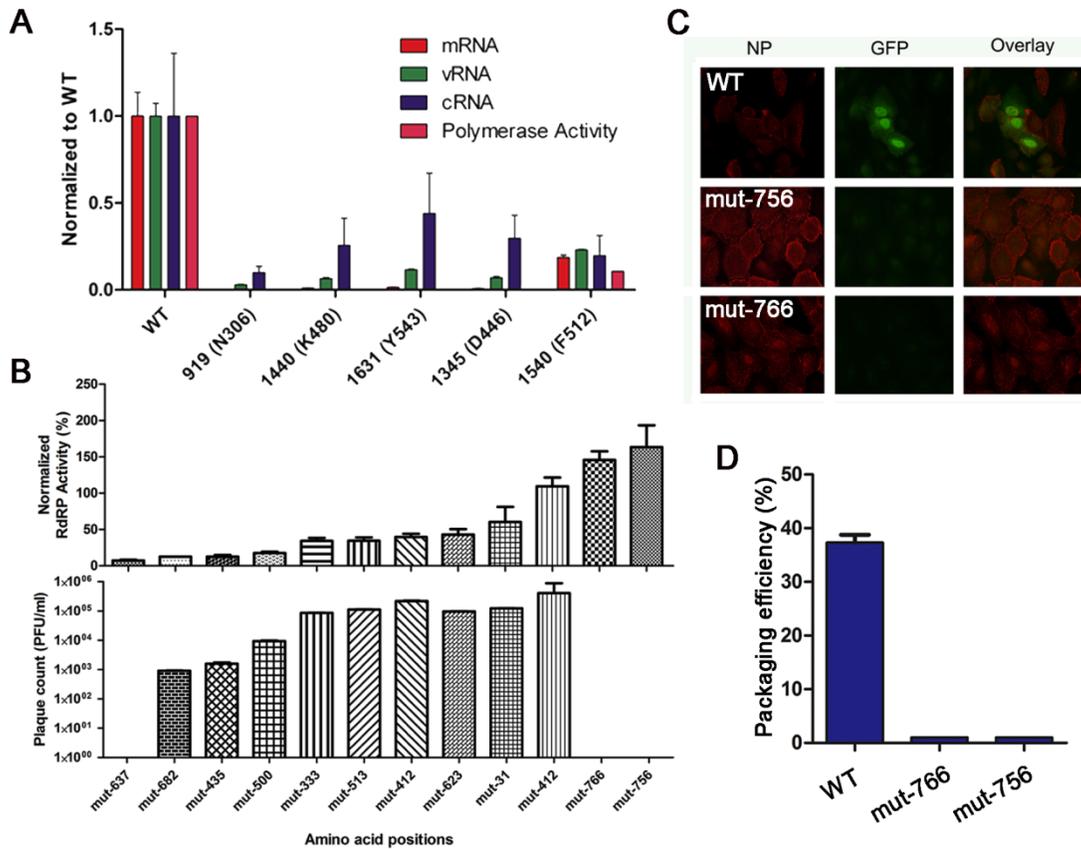
Figure 3. Characterization of PB1 TBC region and viral packaging signals. (A) The effects of individual insertion mutants on the m/c/vRNA synthesis of influenza virus detected using qPCR. **(B)** The polymerase activity and PFU value of individual PB1 mutants. **(C)** Viral packaging assay. **(D)** Packaging signal detection using fluorescent microscopy.

The heterotrimeric interfaces of the influenza polymerase complex

To elucidate the interactions between the subunits of the RdRp, we mapped the viable mutants of PA, PB1 and PB2 recovered from passage 3 onto the polymerase complex protein structure (Fig. 4A). There are several locations where all three subunits share the same interface (Figs. 4C-E). These encompass the entire PA subunit (PAn, PAc-head and PAc-mouth), the N terminus of PB2 (40-256) and the C terminus of PB1 (425-749; Fig. S4). Our mutational studies were able to recover only a handful of viable mutants at these sites. PB1 and the entire PB2-N1 domain (1-250) also share an interface. Functional analysis of individual insertion mutants we

isolated from these domains showed that they lacked any polymerase activity (Fig. S4 and Table S3).
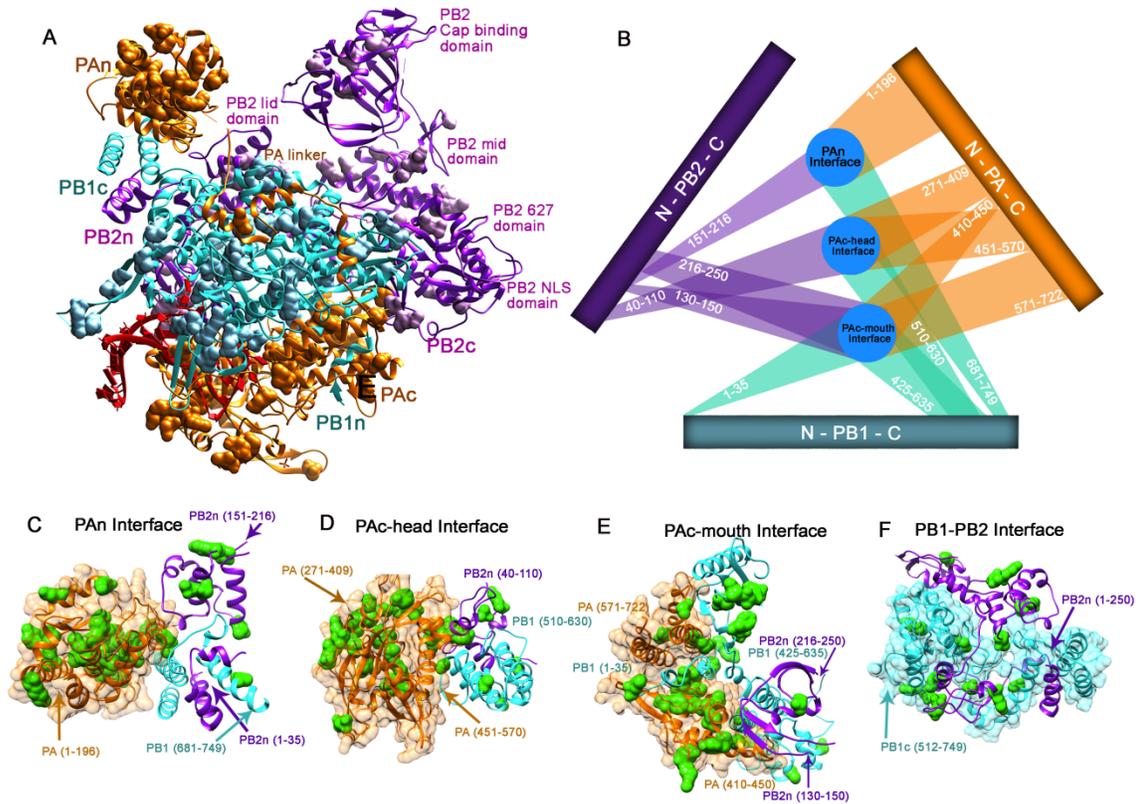
Figure 4: The heterotrimeric interfaces of the influenza polymerase complex. (A) Mapping of all viable mutants onto the crystal structure of the influenza polymerase complex. **(B)** Schematics of the heterotrimer interfaces. **(C-F)** The interface locations labeled onto the polymerase structure. The viable mutants from passage 3 are labeled in green.

Critical role of the inter-domain linkers in the dynamics and functions of influenza polymerase complex

The structure of the influenza RdRP reveals a tightly interlocking architecture between the PA, PB1 and PB2 subunits. The gaps between domains are spanned by numerous interdomain linkers. Using genome-wide profiling, we found that most of the linker regions in all three polymerase proteins do not tolerate any insertional mutations (Fig. 5A), which is unusual given the flexible and non-conserved role typically observed in other proteins with linkers of this length.

We hypothesize that the RdRP's linkers play an indispensable role in achieving multiple polymerase functions.

Although they are spread far apart from each other, the domains of PB2 are connected by an inter-domain linker (highlighted in yellow in Fig. 5A). The N1 linker (33-42) and the N1-Lid linker (100-152) of the PB2 subunit connect the N1 and Lid domains, and are packed closely against PB1c as well as PAn. No viable mutants were discovered in any of these regions, and insertion at nt position 115 completely abolished the polymerase activity (Fig. S4C). The Lid-Mid linker (212-251) and Cap-627 linker (483-538) were highly intolerant of mutation, as was the Mid-Cap linker (316-322), which was proposed to facilitate the conformational change in PB2 (7, 18). None of the single clones found in the linker regions had polymerase activity (Figs. S4A and Table S3). The PB1 subunit is the central subunit of the influenza polymerase complex. We found two distinct linker regions (highlighted in blue in Fig. 5B) at PB1c and PB1n which are mainly involved in binding PB2 and PA, respectively, and are very intolerant to mutation. The PAc and PAn domains are spread far apart and are proximal to different sides of PB1. The PA linker (196-257, highlighted in red in Fig. 5C) that bridges them, of which a portion has been previously identified and studied (30), circles around PB1 (7). Herein, we have also analyzed a region of PB1 (315-345) which directly interacts with the PA linker α7, α8 and α9 helices (201-257). Using randomly selected single clones, we found that insertions in the PB1-PA linker interaction sites greatly reduce polymerase activity (Fig. S4A and Table S2). Furthermore, we highlighted two additional linkers, the extended PA linker (258-290) and the PA arch (366-397), which may interact with PAc and vRNA, respectively. None of the three linkers in the PA subunit are very tolerant to insertion mutations. Linear schematics of the mutants recovered from passage 3 of each library mapped onto the linker positions are depicted in figure 5C. Our data suggest that genome-wide insertion mutagenesis can also be used to highlight key protein-protein interaction domains for further studies.
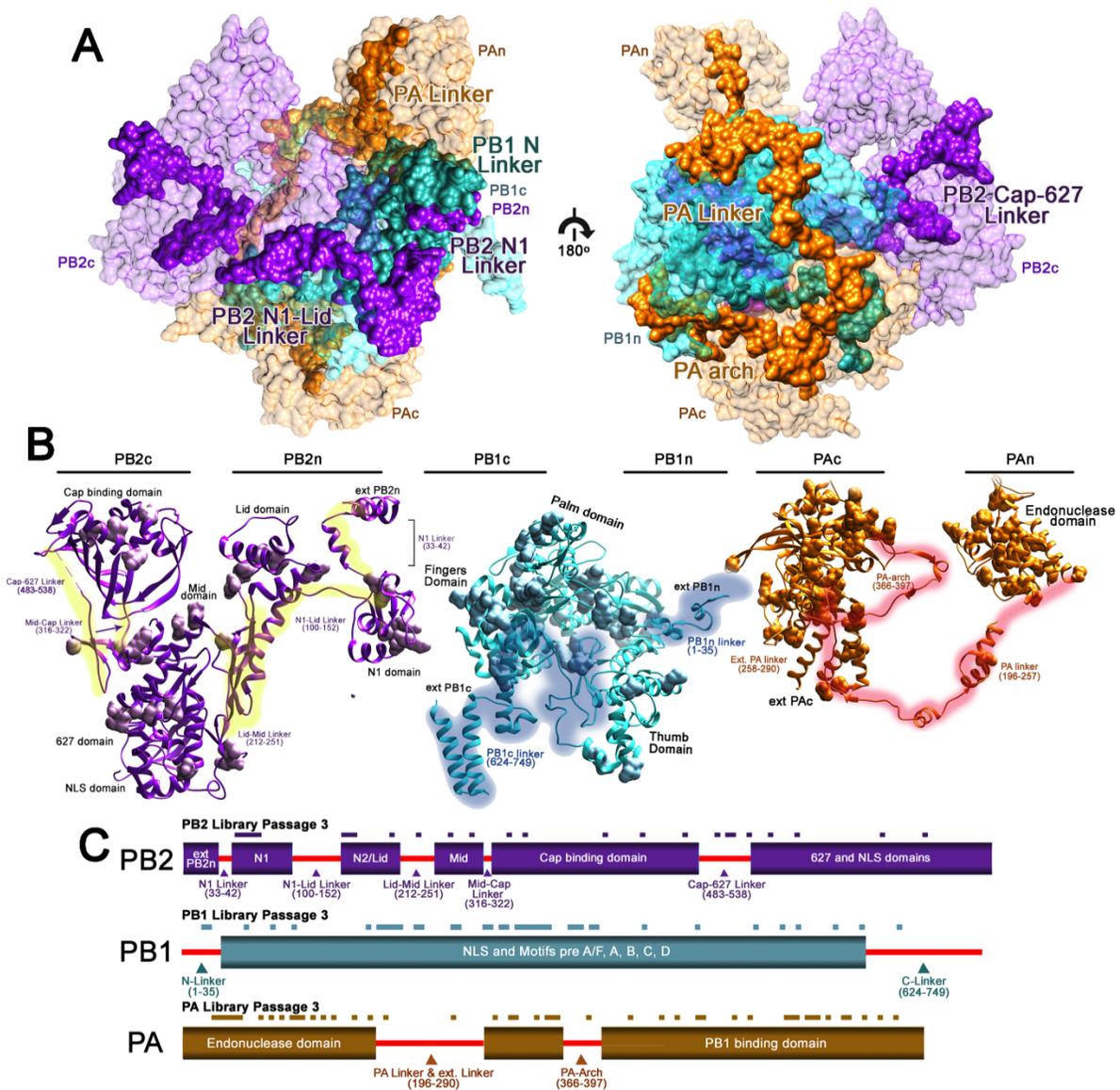
Figure 5: Critical role of the inter-domain linkers in the dynamics and functions of influenza polymerase complex. (A) The locations of each linker from the subunit of the entire RdRP complex are highlighted. **(B)** Each identified linker for PB2, PB1 and PA are highlighted in yellow, cyan and red, respectively. The viable mutants from passage 3 are labeled with surface display. **(C)** Schematics of the linker and viable mutants from passage 3 from each subunit is shown.

Critical ionic interaction between the channel and linker of the PA protein

Genome-wide mutagenesis of the PA subunit confirmed the importance of the PA linker and channel regions, which have been poorly understood. To further clarify why almost all of the

insertion mutants in these regions were nonviable, we generated specific mutations in the PA linker region.

Previous research on the crystal structure of the PAc domain has revealed the presence of a channel-like structure in proximity to the putative RNA-binding groove (24). Although several amino acid residues around this area are known to affect polymerase activity and/or viral growth (6, 31), the biological significance of this channel has yet to be demonstrated and its function remains elusive. We therefore systematically mutated the amino acid residues that line the inner surface of the channel (K461, E524, K536 and E410) as well as those that connect the channel to the RNA-binding groove (R566 and K539; Figure 6A). Alanine substitution of any of these amino acids resulted in a significant reduction of both binding affinity to PB1 and polymerase activity, with E524A showing the most dramatic effect (Figs. 6B and 6C).

We used a bimolecular luminescence complementation assay (BiLC) to examine protein binding affinity. The PA linker fused with the N-terminal half of *Gaussia* luciferase (GN), when expressed together with the full length PB1 protein fused with the C-terminal half, showed strong luciferase activity. We found that a fragment of the PA linker region (196-250; α7-9 helices of PA shown in figure S5) could bind to PB1, but not 220-250 or 220-270 (Figs. 6D and E). We made a series of deletion and insertion mutants at three different locations in the PA linker (amino acids 196, 265 and 285, respectively) to further explore its functions. We found that insertion or deletion of five random amino acids (Ala-Leu-Gly-Ser-Thr) at PA-196 did not affect PB1 binding affinity, but did abolish polymerase activity and RNA synthesis (Figs. 6F, H and I). Deleting or inserting at PA-265 and PA-285 severely impaired its binding affinity to PB1, as well as RdRp polymerase activity. Remarkably, simultaneous insertion and deletion in the original sequence to maintain the length of the linker completely or partially restored binding affinity, polymerase activity, and RNA synthesis. This finding supports our model that the lengths of the linkers are critical for polymerase binding and function.

A short peptide comprised of residues 279-289 in the extended PA linker occupies the channel. The structural model suggests that two positively charged amino acids in this peptide (R279 and K281) can potentially interact with the two negatively charged amino acids on the inner surface of the channel (E410 and E524, respectively). Although alanine substitution in R279 and K289 did not alter the binding affinity to PB1, the polymerase activities of R279A, K281A and K289A were significantly reduced. We confirmed that R279A and K289A retain the heterotrimeric form of the influenza virus polymerase complex using TAP pull down of the influenza virus polymerase complex (Fig. S6). We performed additional mutagenesis studies to further explore the nature of these interactions. It appears that electrostatic interaction is critical, since changing the residues to the opposite charge completely abolished PA function, and neutralizing the negative charges of E410 and E524 through E to Q mutations also severely compromised polymerase activity (Fig. 6C). We also showed that positively charged E410R and E524K mutations abolish the heterotrimeric assembly of the purified influenza virus polymerase proteins (Fig S6). Furthermore, we performed an *in vitro* replication assay using ApG primed transcription and found that ionic interaction is also functionally important for RNA synthesis (Fig. S7). We showed that oppositely charged, but not similarly or uncharged mutations in residues E410, E524, K279 and K281 could completely abolish RNA synthesis (Figs. 6F and 6G). The PA linker co-IP supports the model in which the linker is not only crucial for binding with PB1, but mediates specific interactions with PAc via charged residues (channel region; Figs. 6J and 6K).
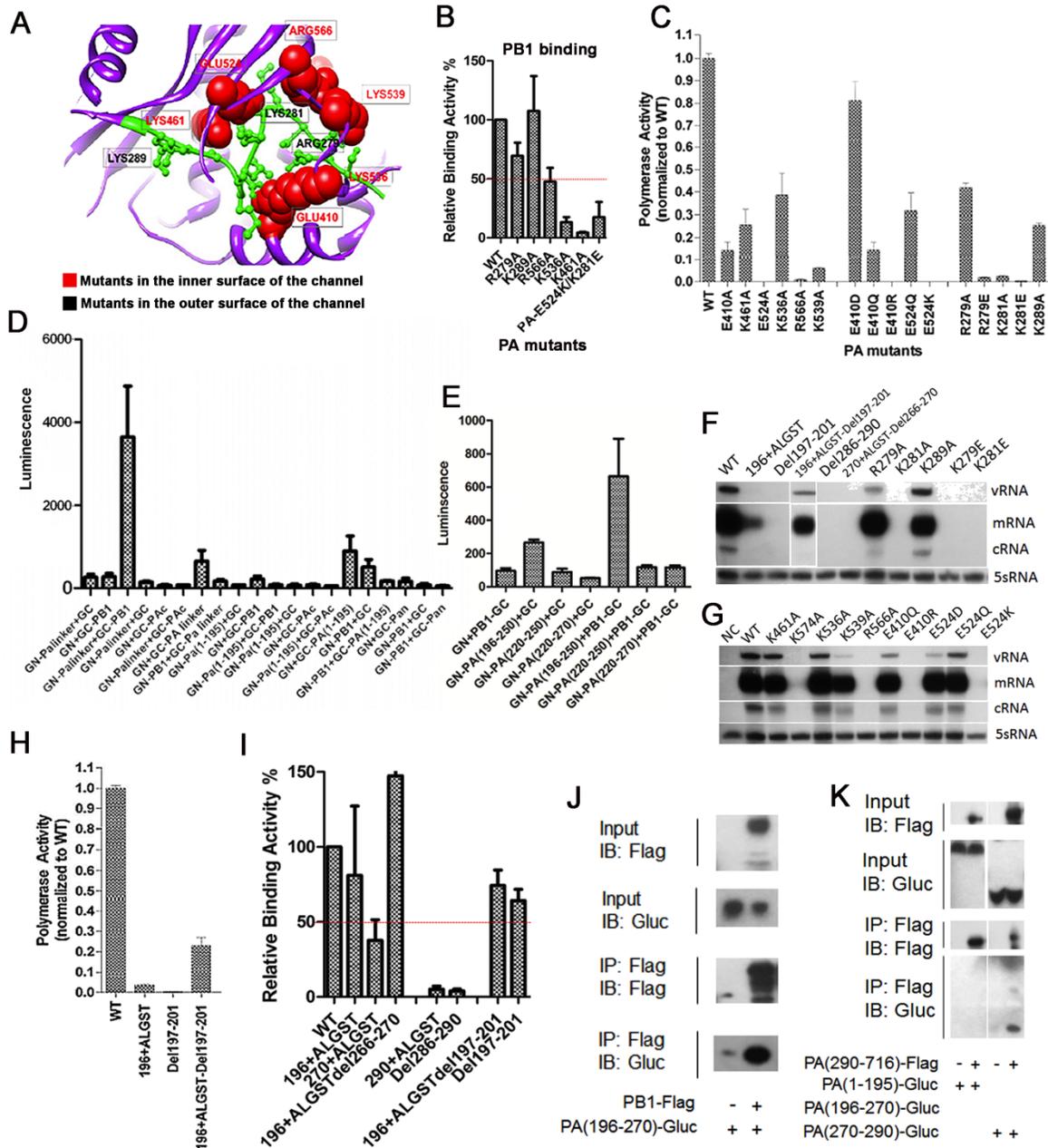
Figure 6: Critical ionic interaction between the linker and channel of PA protein

(A) Residues within the PAc channel region (K461, E524, K536 and E410), as well as those that connect the channel to the RNA-binding groove (R566 and K539), are labeled in the PA protein structure. **(B)** We used BiLC to examine protein binding affinity. Luminescence of *Gaussia luciferase* (GLuc)-fused PA linker was shown for indicated mutants. **(C)** The RdRp activity was measured using the mini-replicon assay. **(D and E)** PA linker truncations were generated to examine their affinity against PB1. Relative binding affinity against PB1 using BiLC was shown.

Additional PA linker mutants were generated. **(F and G)** To analyze the level of m/c/vRNA synthesis in the PA mutants, A/WSN/33 virus-derived PB1, PB2, NP and the indicated PA mutants were co-expressed in 293T cells with pPolI-NA, which can produce NA vRNA. Total RNA was analyzed using a primer extension assay. The relative polymerase activity **(H)** and PB1 binding affinity **(I)** of the PA mutant were shown using the mini-replicon assay and BiLC, respectively. **(J and K)** PB1-PA linker and PA linker-PAc interactions between the truncated PA linker mutants and PA channel region were shown using co-IP.

3.1.c DISCUSSION

Herein, we took an unbiased functional genomics approach to generate a comprehensive functional map of the microdomains of the influenza RdRp heterotrimer, which allow us to rapidly pin-point microdomains of each individual subunit. Guided by this data, we have identified critical domains in the PA linker and channel regions. We showed the critical role these linkers play in coordinating the spatial interactions of the polymerase complex. We also found that the extended PA linker forms ionic interactions with the PA C-terminal channel.

Transposon-based mutagenesis is a valuable and cost-effective tool that can identify critical segments of uncharacterized proteins, guiding further functional analysis (17, 32, 33). We acknowledge that 15nt insertions are large enough to potentially cause significant changes to the structure or stability of the protein, which could impact our assessment of the function of that specific targeted region. Although we do not have the tools to characterize the global conformation changes caused by these insertions, we were able to verify that subunit expression was only rarely affected in the hundreds of mutant clones that we screened, as the vast majority had a protein level comparable to wt samples (data not shown). Given the extensive coverage of our mutant libraries, the loss of a few individual clones within each functional region—which span tens to hundreds of amino acids long—would not be expected to affect our overall observations about those regions. In addition, functional assays of polymerase activity from these clones were consistently in agreement with the viability profile seen in the initial genotyping screen. This confirms that the mutagenesis resulted in loss-of-function of these critical subunits.

Guided by the recently solved crystal structure, we highlighted three heterotrimeric interfaces, and mutations at these locations almost always resulted in non-viable virus with partially or fully abolished polymerase activity. The PB1c and PB2n domains are important for their interactions with each other. However, they may also play a much larger role in the complex as a whole, as they share an interface with the entire PA subunit.

93

A result of our unbiased functional genomics analysis was the surprise discovery of the critical importance of a sophisticated web of linkers connecting the different subunit domains. We identified a total of nine linkers, most of which could not tolerate any insertional mutations and resulted in abolished polymerase activity. We conclude that the linkers play a critical role in the spatial arrangement required to achieve a functional polymerase complex. In addition, our data demonstrates the importance of the linker length, as the insertion of five amino acids at almost any position in the entire linker region led to non-viable virus. We have also shown that the deletion of five amino acids at different positions abolished the polymerase activity, whereas the addition of five amino acids near the deletion site could partially rescue it. Our mutational studies have demonstrated the essential function of the PA linker in supporting polymerase activity, especially when it affects the interactions between PB1 and PA.

Our comprehensive functional map of the influenza polymerase complex will help in the future design and screening of novel anti-influenza drugs. One potential target is the TBC of the PB1 protein, in which we have shown that almost all of the surrounding residues are conserved. Another strategy is to use small molecules to disrupt the ionic interactions between the linker and the channel of the PA protein. The size of the channel could permit small molecules to enter, and we have shown that every ionic interaction is essential for RNA polymerase activity. Finally, targeting the critical PA linker itself could be the means of fully disrupting influenza virus transcription.

MATERIAL AND METHODS

Cells and plasmids

The human embryonic kidney 293 cells (HEK293T) and Madin-Darby canine kidney cells (MDCK) were purchased from ATCC and maintained in Dulbecco's modified Eagle's medium containing 10% Fetal Bovine Serum (FBS; Gibco) and penicillin/streptomycin (100 units/ml and 50 µg/ml, respectively). All cells were incubated at 37°C and 5% CO2. Transfection assays were conducted according to the manual. Plasmids pPoII-CAT-RT and influenza A/WSN/33 virus-derived plasmids pcDNA-PB2/PB1/PA/NP were described previously (6).

Generation of Mutagenesis Library of PA, PB1 and PB2

All three gene segments (PA, PB1 and PB2) of influenza virus A/WSN/33 underwent Mu-transposon mediated mutagenesis (MGS kit, Finnzymes) according to the manufacturer's instructions. This resulted in three mutant libraries, with each gene containing a randomly positioned 15nt insertion that includes the 10nt unique sequence 5'-TGCGGCCGCA-3'. The mutated segment and seven wild-type plasmids were co-transfected into HEK293T cells for virus packaging (34), and virus was then collected and used to infect MDCK cells for 24 to 48 hours for each of the three consecutive passages. The 2nd and 3rd passage in MDCK were achieved through subsequent inoculation of 0.1 M.O.I. of the viral pool from the previous passage. The 0.1 M.O.I still maintains genetic diversity and doesn't introduce artificial stochastic bottlenecks. To ensure that there was no contamination with DNA used for transfection, the media was replaced 12 hours post-infection. Total RNA of the mutant pool from all passages of virus was isolated using TRIzol and converted to cDNA with iScript (Life Technology). The mutants were genotyped by PCR amplification with gene-specific forward primers and a Vic-labeled reverse primer against the 10nt insertion. To minimize PCR bias, we designed forward primers to hybridize 300 to 500nt apart and ensure complete coverage of each gene (Table S1). The final product was sequenced using a 96-capillary DNA Analyzer with the size standard Liz 500 (3730xl DNA Analyzer, Applied Biosystems) at the UCLA GenoSeq Core facility. Sequencing data was analyzed for clarity using

the ABI software, with the following criteria: (1) All data passed the standard default detection level; (2) The first 70bp were removed due to non-specific background noise; (3) All data were aligned to the nearest bp or amino acid position in the specific gene; (4) All genotyping experimental data were normalized with WT WSN, non-transfected cells and a different gene library as controls. This eliminated non-specific data from the PCR, primers and the DNA Analyzer.

Viral packaging assay

A packaging reporter was constructed by replacing nt 67-2158 of the PB1 coding sequence (in the PB1 plasmid from the 8-plasmid reverse genetics system) with the open reading frame of GFP. The resulting construct contains GFP flanked by the minimal signal for efficient packaging of the PB1 segment (i.e. the UTRs plus 66nt and 50nt of the PB1 coding sequence at the 3' and 5' ends, respectively) (35). Mutants containing 15bp insertions at the indicated positions were generated by site-directed mutagenesis. For the viral packaging assay, 293T cells were transfected with the 8-plasmid system plus the packaging reporter. 48hrs later, supernatants from transfected cells were collected, cleared by centrifugation, and used for infection on fresh target cells. At 16hrs post-infection, cells were stained with an anti-NP antibody (AA5H) to quantify total infection. A GFP+NP+ double positive population indicates infection by viruses that have packaged the GFP packaging reporter. Packaging efficiency was calculated by dividing the percentage of GFP+NP+ double positive cells by the percentage of NP+ cells.

Construction of the polymerase complex model

The influenza polymerase complex functional map was constructed using the published crystal structure of the complete polymerase complex (7). We adapted the influenza A polymerase complex structure (PDB ID: 4WSB), and all mappings were made using UCSF Chimera (36).

Primer extension assay

Plasmids pPolI-CAT-RT and A/WSN/33 virus-derived plasmids pcDNA-PB2/PB1/PA/NP were described previously.

Mini replicon system for influenza A virus

HEK293T cells were transfected with Lipo2000 (Invitrogen) in 6-well plates. A/WSN/33-derived PB2, PB1, PA, NP-expressing plasmids and a luciferase reporter plasmid pPolI-NP-luc were co-transfected into 293T cells. Renilla luciferase activity was used for normalizing variation in transfection efficiency. 24 hours post-transfection, firefly luciferase and renilla luciferase of cell lysates were detected by the Dual-Glu Luciferase Assay system (Promega) using a microplate reader (Tecan,GENiosPlus). The activity detected with samples containing WT PA was set to 100%.

Polymerase assay and strand-specific real-time PCR

A minigenome reporter for influenza virus polymerase activity was constructed. Briefly, the *Gaussia* luciferase (GLuc) gene flanked by the UTRs from the PA segment of the WSN virus was inserted between the human pol I promoter and mouse pol I terminator in the reverse-sense orientation. Co-transfection of WSN PB1, PB2, PA and NP expression constructs results in the expression of GLuc protein, which was used as an indicator of polymerase activity. The mRNA, vRNA and cRNA transcript levels were quantified using a strand-specific real-time RT-PCR method developed by Kawakami *et al*, which has been shown to be able to distinguish between the three different types of RNA with high specificity. The same tags as described by Kawakami *et al* were used, and primers were designed to target the GLuc gene (37).

BiLC assay

The GLuc gene was divided at site 109 into C and N terminal fragments labeled GC and GN, respectively. PB1 was cloned upstream of GN and PA was cloned upstream of GC to create

fusion products. Additional constructs containing truncated fragments of PA (residues 1-195, 196-250, 196-270, 220-250, 220-270), or with a 5 amino acid addition (ALGST) at position 196 or 270, with a 5 amino acid deletion from position 266 to 270 were also created. The fused proteins were co-expressed in 293T cells for 24 hours. Cells were lysed and protein interactions were assessed with the *Gaussia* luciferase kit (Promega) using a microplate reader.

Protein immunoblot

PB1 and PA were detected by rabbit polyclonal serum (1:1500) and alpha-Tubulin was detected by anti-Tubulin antibody (1:2000). Blots were incubated with primary antibodies for 2 hours at room temperature, then incubated for 1h at RT with HRP-conjugated secondary antibodies (1:10000,Beijing ZSBio).

TAP pull down of influenza virus polymerase complex

Recombinant PB2-TAP tagged influenza virus polymerase containing PA and PB1 were purified, then separated by SDS-PAGE and silver staining as described previously (22), Amino Acid Residues in the N-Terminal Region of the PA Subunit of Influenza A Virus RNA Polymerase Play a Critical Role in Protein Stability, Endonuclease Activity, Cap Binding, and Virion RNA Promoter Binding).

ApG-primed transcription

Reactions were conducted with recombinant heterotrimer polymerase and a model vRNA promoter consist by short 15 and 14 nt synthetic 5' and 3' ends in vitro (37℃, 1H) as described previously (22).

REFERENCE

1. WHO. 2003. Influenza: Fact sheets.

2. Kolpashchikov DM, Honda A, Ishihama A. 2004. Structure-function relationship of the influenza virus RNA polymerase: primer-binding site on the PB1 subunit. Biochemistry 43:5882-5887.

3. Kabsch W, Sander C. 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22:2577-2637.

4. Buchan DW, Ward SM, Lobley AE, Nugent TC, Bryson K, Jones DT. 2010. Protein annotation and modelling servers at University College London. Nucleic Acids Res 38:W563-568.

5. Bao Y, Bolotov P, Dernovoy D, Kiryutin B, Zaslavsky L, Tatusova T, Ostell J, Lipman D. 2008. The influenza virus resource at the National Center for Biotechnology Information. J Virol 82:596-601.

6. Fodor E, Crow M, Mingay LJ, Deng T, Sharps J, Fechter P, Brownlee GG. 2002. A single amino acid mutation in the PA subunit of the influenza virus RNA polymerase inhibits endonucleolytic cleavage of capped RNAs. J Virol 76:8989-9001.

7. Pflug A, Guilligay D, Reich S, Cusack S. 2014. Structure of influenza A polymerase bound to the viral RNA promoter. Nature.

8. Fodor E. 2013. The RNA polymerase of influenza a virus: mechanisms of viral transcription and replication. Acta Virol 57:113-122.

9.   Guilligay D, Tarendeau F, Resa-Infante P, Coloma R, Crepin T, Sehr P, Lewis J, Ruigrok RW, Ortin J, Hart DJ, Cusack S. 2008. The structural basis for cap binding by influenza virus polymerase subunit PB2. Nat Struct Mol Biol 15:500-506.

10.   Dias A, Bouvier D, Crepin T, McCarthy AA, Hart DJ, Baudin F, Cusack S, Ruigrok RW. 2009. The cap-snatching endonuclease of influenza virus polymerase resides in the PA subunit. Nature 458:914-918.

11.   Yuan P, Bartlam M, Lou Z, Chen S, Zhou J, He X, Lv Z, Ge R, Li X, Deng T, Fodor E, Rao Z, Liu Y. 2009. Crystal structure of an avian influenza polymerase PA(N) reveals an endonuclease active site. Nature 458:909-913.

12.   Pritlove DC, Poon LL, Devenish LJ, Leahy MB, Brownlee GG. 1999. A hairpin loop at the 5' end of influenza A virus virion RNA is required for synthesis of poly(A)+ mRNA in vitro. J Virol 73:2109-2114.

13.   Gastaminza P, Perales B, Falcon AM, Ortin J. 2003. Mutations in the N-terminal region of influenza virus PB2 protein affect virus RNA replication but not transcription. J Virol 77:5098-5108.

14.   Hatakeyama D, Shoji M, Yamayoshi S, Hirota T, Nagae M, Yanagisawa S, Nakano M, Ohmi N, Noda T, Kawaoka Y, Kuzuhara T. 2014. A novel functional site in the PB2 subunit of influenza A virus essential for acetyl-CoA interaction, RNA polymerase activity, and viral replication. J Biol Chem 289:24980-24994.

15.   Tarendeau F, Boudet J, Guilligay D, Mas PJ, Bougault CM, Boulo S, Baudin F, Ruigrok RW, Daigle N, Ellenberg J, Cusack S, Simorre JP, Hart DJ. 2007. Structure and nuclear import function of the C-terminal domain of influenza virus polymerase PB2 subunit. Nat Struct Mol Biol 14:229-233.

16.   Resa-Infante P, Jorba N, Zamarreno N, Fernandez Y, Juarez S, Ortin J. 2008. The host-dependent interaction of alpha-importins with influenza PB2 polymerase subunit is required for virus RNA replication. PLoS One 3:e3904.

17.   Arumugaswami V, Remenyi R, Kanagavel V, Sue EY, Ngoc Ho T, Liu C, Fontanes V, Dasgupta A, Sun R. 2008. High-resolution functional profiling of hepatitis C virus genome. PLoS Pathog 4:e1000182.

18.   Reich S, Guilligay D, Pflug A, Malet H, Berger I, Crepin T, Hart D, Lunardi T, Nanao M, Ruigrok RW, Cusack S. 2014. Structural insight into cap-snatching and RNA synthesis by influenza polymerase. Nature.

19.   Gonzalez S, Ortin J. 1999. Characterization of influenza virus PB1 protein binding to viral RNA: two separate regions of the protein contribute to the interaction domain. J Virol 73:631-637.

20.   Li ML, Ramirez BC, Krug RM. 1998. RNA-dependent activation of primer RNA production by influenza virus polymerase: different regions of the same protein subunit constitute the two required RNA-binding sites. EMBO J 17:5844-5852.

21.   Biswas SK, Nayak DP. 1994. Mutational analysis of the conserved motifs of influenza A virus polymerase basic protein 1. J Virol 68:1819-1826.

22.   Hara K, Schmidt FI, Crow M, Brownlee GG. 2006. Amino acid residues in the N-terminal region of the PA subunit of influenza A virus RNA polymerase play a critical role in protein stability, endonuclease activity, cap binding, and virion RNA promoter binding. J Virol 80:7789-7798.

23.   Hiromoto Y, Saito T, Lindstrom SE, Li Y, Nerome R, Sugita S, Shinjoh M, Nerome K. 2000. Phylogenetic analysis of the three polymerase genes (PB1, PB2 and PA) of influenza B virus. J Gen Virol 81:929-937.

24.   He X, Zhou J, Bartlam M, Zhang R, Ma J, Lou Z, Li X, Li J, Joachimiak A, Zeng Z, Ge R, Rao Z, Liu Y. 2008. Crystal structure of the polymerase PA(C)-PB1(N) complex from an avian influenza H5N1 virus. Nature 454:1123-1126.

25.   Poch O, Sauvaget I, Delarue M, Tordo N. 1989. Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. EMBO J 8:3867-3874.

26.   Chunfeng lA, W.; Yousong, P.; Jingfeng, W.;Yang, G.; Zhigao, C.; Hong, Z.; Yongqiang, W.; Jiuhong, D.; Lulan, W.; F. Xiao-Feng, Q.; Genhong, C.; Tao D. 2014. Integrating computational modeling and functional assays to decipher the structure-function relationship of influenza virus PB1 protein. Scientific Report In press.

27.   Ferrer-Orta C, Arias A, Escarmis C, Verdaguer N. 2006. A comparison of viral RNA-dependent RNA polymerases. Current opinion in structural biology 16:27-34.

28.   Marsh GA, Rabadan R, Levine AJ, Palese P. 2008. Highly conserved regions of influenza a virus polymerase gene segments are critical for efficient viral RNA packaging. J Virol 82:2295-2304.

29.   Hutchinson EC, von Kirchbach JC, Gog JR, Digard P. 2010. Genome packaging in influenza A virus. J Gen Virol 91:313-328.

30.   Guu TS, Dong L, Wittung-Stafshede P, Tao YJ. 2008. Mapping the domain structure of the influenza A virus polymerase acidic protein (PA) and its interaction with the basic protein 1 (PB1) subunit. Virology 379:135-142.

31.   Regan JF, Liang Y, Parslow TG. 2006. Defective assembly of influenza A virus due to a mutation in the polymerase subunit PA. J Virol 80:252-261.

32.   Heaton NS, Sachs D, Chen CJ, Hai R, Palese P. 2013. Genome-wide mutagenesis of influenza virus reveals unique plasticity of the hemagglutinin and NS1 proteins. Proc Natl Acad Sci U S A 110:20248-20253.

33.   Wu NC, Young AP, Al-Mawsawi LQ, Olson CA, Feng J, Qi H, Chen SH, Lu IH, Lin CY, Chin RG, Luan HH, Nguyen N, Nelson SF, Li X, Wu TT, Sun R. 2014. High-throughput profiling of influenza A virus hemagglutinin gene at single-nucleotide resolution. Scientific reports 4:4942.

34.   Hoffmann E, Neumann G, Kawaoka Y, Hobom G, Webster RG. 2000. A DNA transfection system for generation of influenza A virus from eight plasmids. Proc Natl Acad Sci U S A 97:6108-6113.

35.   Liang Y, Hong Y, Parslow TG. 2005. cis-Acting packaging signals in the influenza virus PB1, PB2, and PA genomic RNA segments. J Virol 79:10348-10355.

36.   Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF Chimera--a visualization system for exploratory research and analysis. Journal of computational chemistry 25:1605-1612.

37.   Kawakami E, Watanabe T, Fujii K, Goto H, Watanabe S, Noda T, Kawaoka Y. 2011. Strand-specific real-time RT-PCR for distinguishing influenza vRNA, cRNA, and mRNA. J Virol Methods 173:1-6.
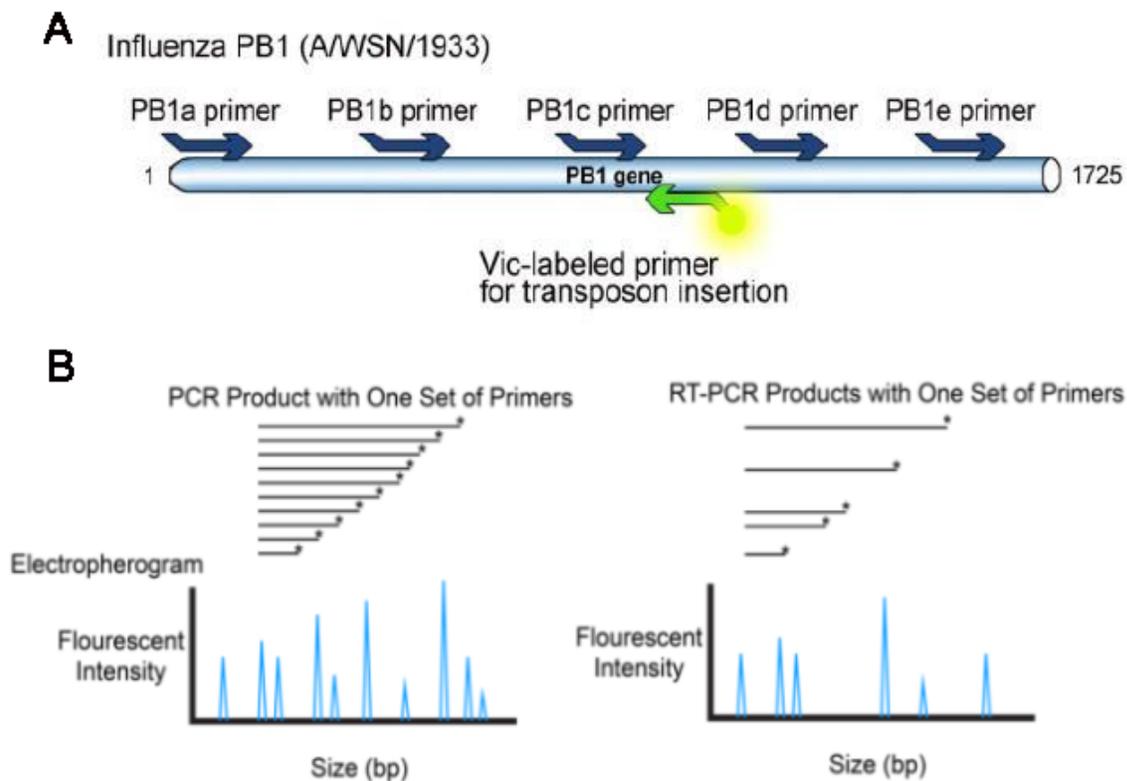
Figure S1. Genome-wide virus mutagenesis and data collection. (A) Individual mutations are genotyped by PCR amplification with a gene-specific primer and a Vic-labeled primer against the 15nt insertion sequence. (B) The mutant pool library can be put through selection such as growth *in vitro* or *in vivo*. The selected pool is collected and genotyped. Essential and non-essential regions in the genome required for growth can be determined by comparing genotyping data from the unselected pool (total mutant library) and selected pool.
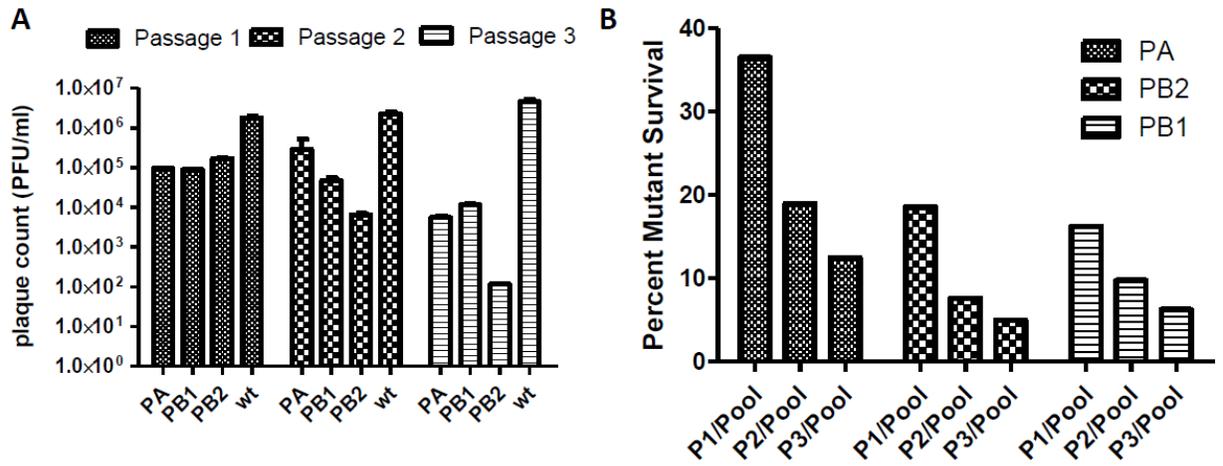
Figure S2. Characterization of virus pool from each passage. (A) The viral titer of WT and mutant pools at the indicated passage was measured by plaque assay using MDCK cells. (B) Total numbers of mutants detected in the initial PB1 mutant DNA library and recovered from the first three passages in MDCK cells after reconstitution of the influenza mutant pool.
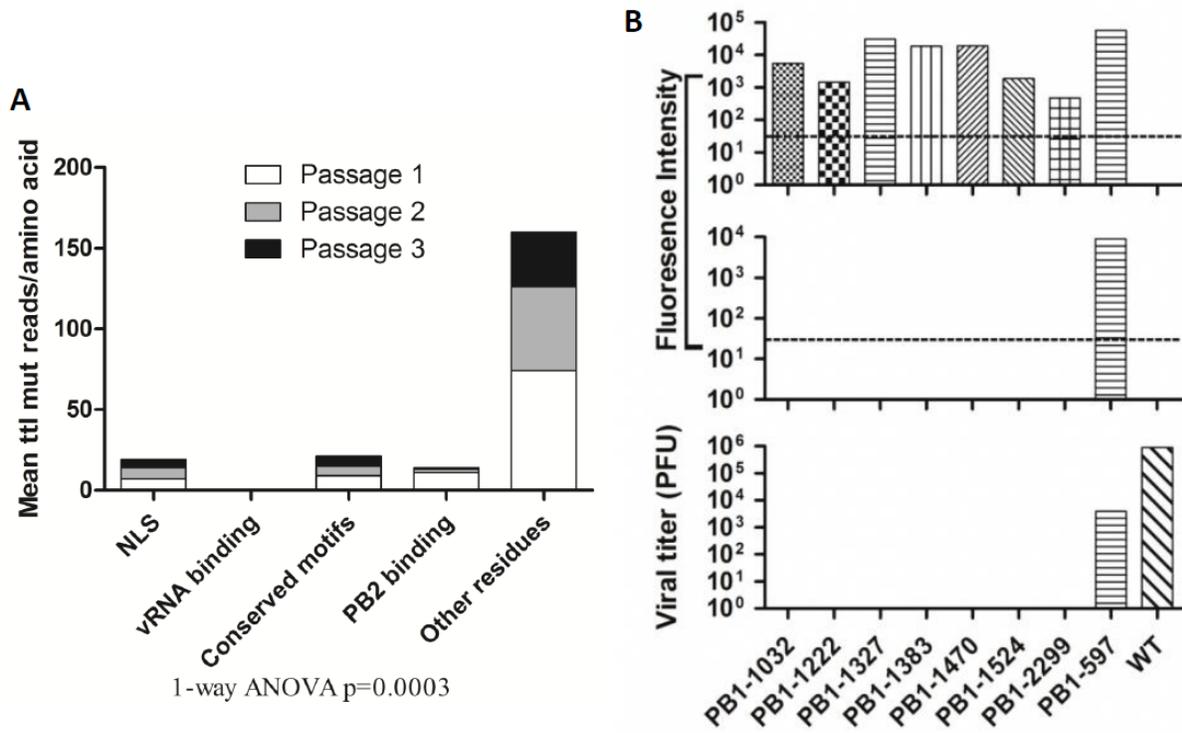
Figure S3. Validation of genotyping data using isolated single clones. (A) The location of significant viable mutants at functional regions of the PB1 gene by amino acid position. NLS, Nuclear Localization Signal. (B) To verify our mutagenesis, the mutant PB1 plasmid library was co-transfected with 7 WT plasmids in 293T cells to recover virus in passage 1. Eight randomly selected single clone mutants were isolated from the plasmid pool and were co-transfected independently with 7 WT plasmids. The graph indicates where single clone mutants and viruses from passage 1 were unable to grow.
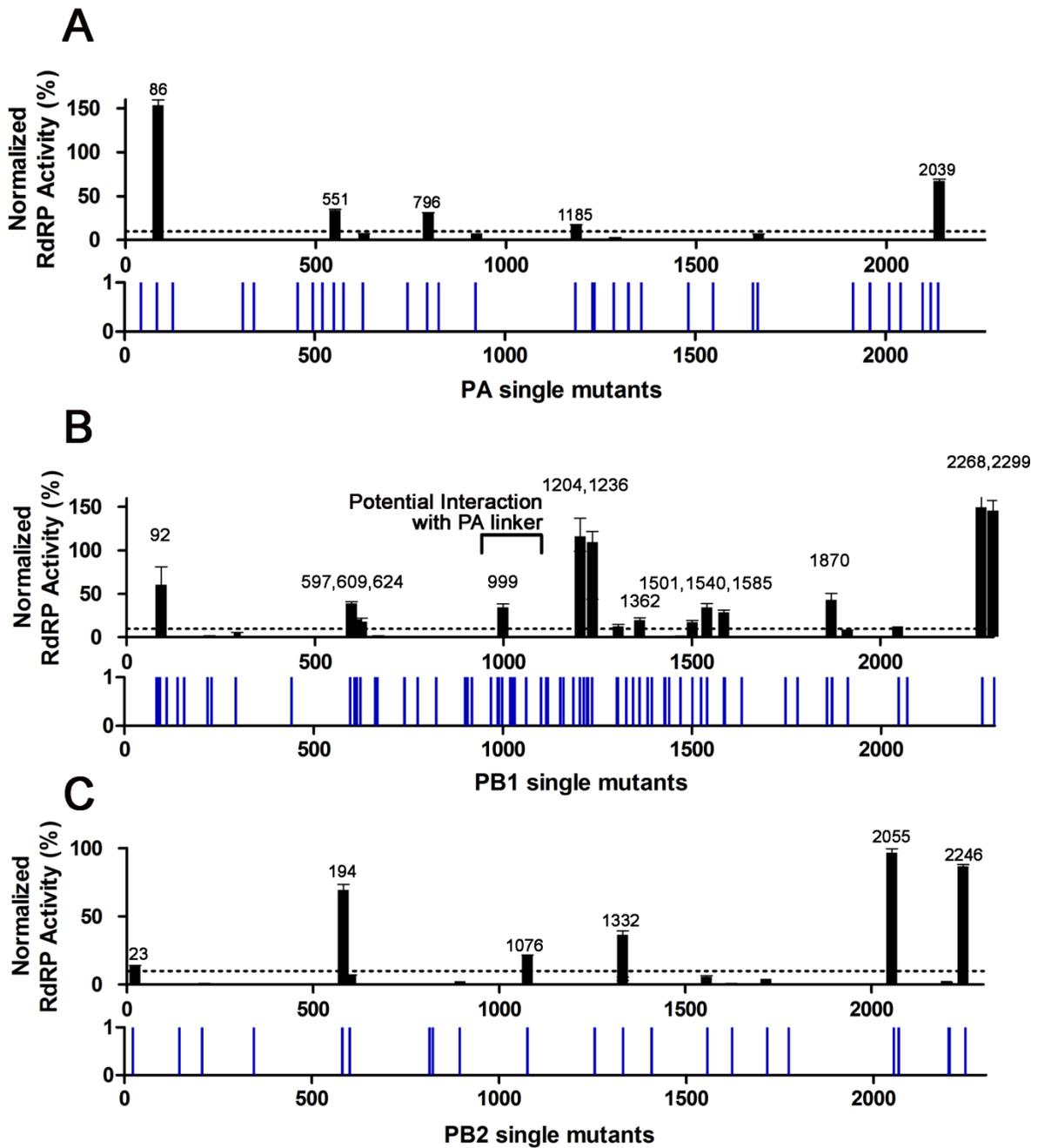
Figure S4. Location and RdRP activity of all isolated single clones from PA, PB1 and PB1 library. The coverage and location of all single mutants from the PA (A), PB1 (B) and PB2 (C) libraries shown in blue bars, whereas the relative polymerase activities are labeled in black bars. **Figure S5**.
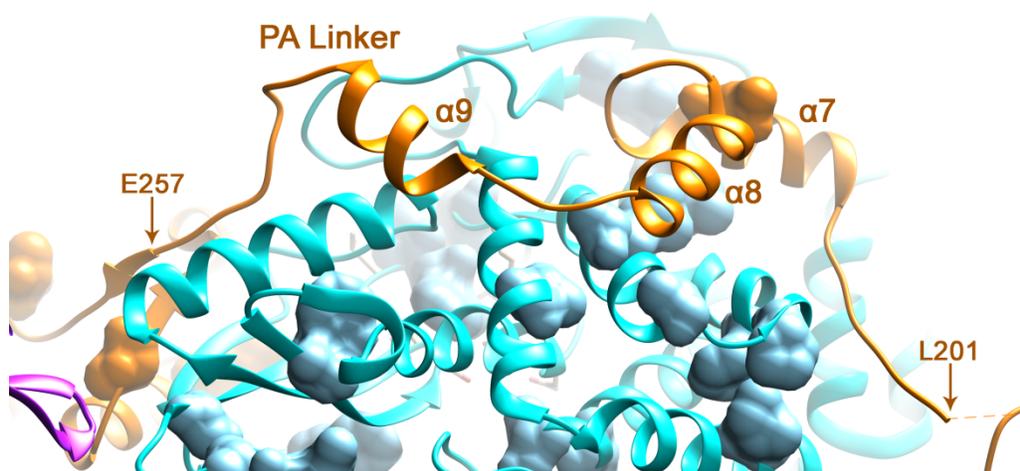
Figure S5. **Structure of the PA linker.** The PA linker that interacts with PB1 subunit is labeled in orange, whereas the PB1 interface is labeled in cyan. PB2 subunit is labeled in purple. Site 264 (green) was highlighted on the PA linker (blue). A charged amino acid tunnel surrounding the linker on PAc is shown in purple. PB1c and PAc are shown as transparent grey and orange surfaces, respectively.
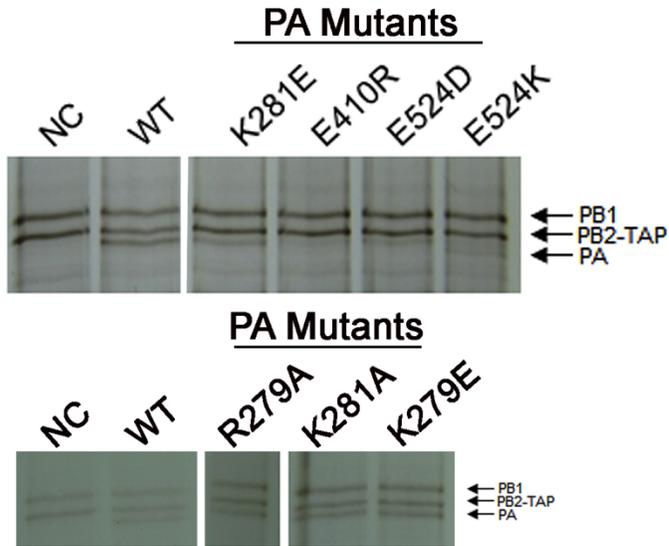
Figure S6. **TAP pull-down of influenza virus polymerase complex with PA mutants.** PB2-TAP tagged heterotrimer polymerase complex was purified in 293T cells at 40 hours after transfection. The positions of PA, PB-TAP and PB1 are indicated.

PA mutants

ApG primed
Transcription
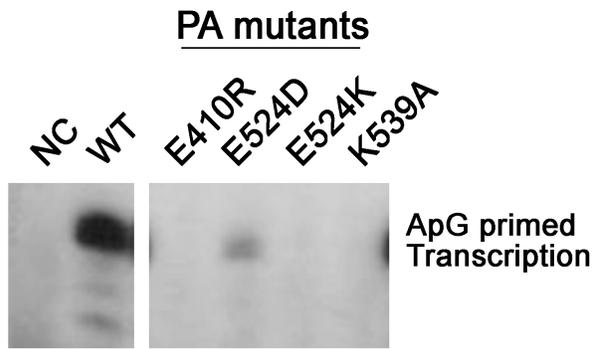
Figure S7. ApG primed transcription assay. *In vitro* ApG primed transcription assay of influenza virus polymerase with WT and indicated mutations on PA. The positions of transcription products are indicated.

PA	Primer Positions & Sequence
13	ATTTGAATGGATGTCAATCC
456	ACACAATAGAAGTGTTCAG
847	GATAAGGAAGATGATGACC
1276	AATCTTGGACAAAAGAGAC
1687	ATACAACATTAGAAATCTCC

PB1	Primer Positions & Sequence
1	GAAGATTTTGTGCGACAATGC
439	ATTCACTGGGGAGGAAATGG
861	TCCAAATTCCTGCTGATGG
1278	ATAGAGCTCGATGAGATTGG
1799	CCTATGTTCTTGTATGTGAGG

PB2	Primer Positions & Sequence
85	ATGTCGCAGTCTCGCACTC
383	AACCTTTGGCCCTGTCCATTTTAG
739	GAGGGGAGGCGAGGAATG
1094	GGTTGGGAGAAGAGCAACAGC
1449	GGGGTAGATGAGTATTCCAGCG
1876	CAAAGCAAAGTGGAATGCAGTTCTC

Figure S8. Primers, nucleotide locations, and sequences.

3.2 GENERATION OF A LIVE ATTENUATED INFLUENZA VACCINE THAT ELICITS BROAD PROTECTION IN MICE AND FERRETS

ABSTRACT

New influenza vaccines that provide effective and broad protection are desperately needed. Live attenuated viruses are attractive vaccine candidates because they can elicit both humoral and cellular immune responses. However recent formulations of live attenuated influenza vaccines (LAIVs) have not been protective. We combined high-coverage transposon mutagenesis of influenza virus with a rapid high-throughput screening for attenuation to generate W7-791, a live attenuated mutant virus strain. W7-791 produced only a transient asymptomatic infection in adult and neonatal mice even at doses 100-fold higher than the LD_{50} of the parent strain. A single administration of W7-791 conferred full protection to mice against lethal challenge with H1N1, H3N2 and H5N1 strains, and improved viral clearance in ferrets. Adoptive transfer of T cells from W7-791-immunized mice conferred heterologous protection, indicating a role for T cell-mediated immunity. These studies present a LAIV development strategy to rapidly generate and screen entire libraries of viral clones.

3.2.a INTRODUCTION

Influenza A virus is a major public health problem. In a typical year, influenza infects as many as 500 million people worldwide and leads to more than 500,000 deaths. In the United States, 5-10% of the population is infected by influenza virus in an average season, resulting in ~220,000 hospitalizations and ~36,000 deaths (WHO, 2003). However, significant mutations in the virus will bypass host immunity to previously exposed strains, leading to considerably greater mortality. The devastating "Spanish flu" pandemic of 1918 was one such example, where the virus itself or complications from secondary infections killed an estimated 40-50 million people worldwide. Vaccination has been the most effective way to prevent the spread of influenza and its complications. Live attenuated influenza vaccines (LAIVs) are known to be more immunogenic than inactivated vaccines, likely because they stimulate both humoral and cell-mediated immune responses (Belshe et al., 2007). Traditionally, attenuated vaccines such as the measles-mumps-rubella and influenza vaccines were made via a forward genetics approach, using random mutagenesis followed by rounds of selection in non-physiological conditions, a time-intensive process that produces few vaccine candidates (Lamb et al., 1981). Moreover, these vaccines need to be reformulated annually due to antigenic drift and poor cross-protection against emerging pandemic strains (Jang and Seong, 2012; Krammer and Palese, 2015). In addition to difficulties in producing and formulating the vaccine, recent findings suggest that the quadrivalent LAIVs used over the last three years have not been protective. The CDC has actually recommended against using the LAIV during the 2016-17 season (Grohskopf et al., 2016). In a proof-of-principle study, we sought to use the reverse genetics system of influenza virus and a transposon mutagenesis system as tools towards the rapid, high-throughput generation and screening of viral clones with attenuated growth *in vivo* as candidate LAIVs.

3.2.b RESULTS

Generation of LAV W7-791 using genome-wide mutagenesis and *in vivo* profiling.

In contrast to the antigenic variability of the influenza hemagglutinin and neuraminidase genes, the matrix (M) gene segment encoding viral capsid protein (M1) and the proton-selective channel protein (M2) has evolved very slowly in all lineages of the virus, and has been evaluated as a target for universal vaccine development (De Filette et al., 2008; De Filette et al., 2005; Huleatt et al., 2008; Ilyinskii et al., 2008; Krammer and Palese, 2015; Neirynck et al., 1999; Schnell and Chou, 2008; Schotsaert et al., 2009; Tompkins et al., 2007; Wang et al., 2014). We created a mutant library comprised of $>10^5$ random insertions of a 15 nucleotide sequence, 5'-NNNNNTGCGGCCGCA-3', in the M gene of influenza A/WSN/1933 H1N1 (WSN), using a Mu-transposon mutagenesis method previously described (Arumugaswami et al., 2008). We then used the reverse genetics system of influenza to transfect the M gene mutant library with the 7 complementary segments of wild-type (wt) WSN into HEK293T cells, creating a mutant virus library (Figures S1A-C). A primer based on the unique sequence contained within the insert was paired with different downstream primers along the length of the gene, generating a genotyping product of a specific length that could be used to extrapolate the position of the insertion. In this way, the mutation coverage of the entire viral pool could be visualized simultaneously. Analysis of this high-throughput genotyping data following *in vitro* infection of the mutant virus library (Figure S1D) shows that the majority of viable mutants (green peaks) are located in the non-conserved regions of the M1 and M2 proteins (Figures S1E and S1F).

In order to search for a potential LAIV, we monitored the *in vivo* growth profiles of the entire mutant virus population. Eight mice were given an intratracheal (*i.t.*) injection of the M gene mutant virus library, and cDNA was synthesized from the RNA recovered from lung homogenates collected 2, 4, 6, and 8 days post-infection (dpi). Through genotyping, we identified the location of each insertion in the pool of surviving mutant viruses at each time point. We observed three distinct growth profiles in these viruses (Figure 1A), possibly related to the mutations' effect on

114

the replication, fitness, or host immune-related properties of the virus. Cluster A* represents an example of the fast-growing and likely disease-causing population. Cluster B* represents a slow-growing population, either due to intrinsically slow growth or suppression due to the host's immune response. By contrast, the viruses in cluster C* grew rapidly during the first 6 days, but were cleared between day 6-8. The advantages of this population are that these candidates persist long enough to trigger strong immune responses, but are too attenuated to cause significant disease in the host. We isolated 67 single mutant clones to screen as potential vaccine candidates (Figures S2A and S2B). We amplified 3 mutants in cluster C* (W7-757, W7-791, W7-797) in MDCK cells and found that W7-791 grew to a high titer (Figure 1B) with a slightly lower cell toxicity (Figure S2C), causing less cell death as measured by LDH viability assay compared to the wt WSN virus. The mutant clone W7-791 had the insertion RHCGRI after the 26th amino acid of the M2 gene (Figure S2D), which is located on the cytoplasmic portion of the proton channel (Figure 1C)(Schnell and Chou, 2008). To ensure that W7-791 does not revert back to the wt WSN strain, we passaged W7-791 *in vitro* for multiple generations and found that the insertion remains unchanged in the genome (data not shown). To examine the *in vivo* stability of W7-791, we passaged the virus consecutively to groups of naïve mice at day 4 dpi (Figures S2E-G). We found that the W7-791 titer decreased after each passage, suggesting that the insertion mutation strain remained attenuated. Intranasal (*i.n.*) immunization of 6-8 week-old mice with different titrations of W7-791 showed no signs of weight loss even at 10^7 $TCID_{50}$ per mouse, while significant weight loss was observed in mice infected with 10^3 $TCID_{50}$ of WSN (Figures 1D and 1E) or 10^4 $TCID_{50}$ of the H3 strain (Figure S3A). Similarly, the viral load in the lungs of W7-791-infected mice measured at 6 dpi was nearly 100-fold lower than in the WSN-infected group (Figures 1F and 1G) or the H3-infected group (Figure S3B). Lung samples collected at 4 dpi in both PBS and W7-791 groups showed no obvious pathology, whereas WT-WSN mice, even at a much lower dose, showed severe tissue damage (Figure S3C). We further tested the safety of W7-791 in 15-day-old neonatal BALB/c mice by *i.n.* injection (10^6, 10^7 or 10^8 $TCID_{50}$) of W7-791 or 10^4 $TCID_{50}$ of WSN. The weight loss (Figure 1H) and lung findings (Figure S3D) suggest that

115

W7-791 also exhibits significantly less pathology than wt WSN in neonatal mice. To test the extent of this particular insertion in causing viral attenuation, we inserted the same mutation into the matrix gene of influenza A/Puerto Rico/8 H1N1 (PR8), another mouse adapted but heterologous H1 strain. We found that this mutant showed the same level of attenuation compared to its parent strain *in vivo* (Figure S4A). Taken together, we have established that the mutant W7-791 influenza strain is sufficiently attenuated to cause only a temporary infection, and is safe at high titers in both adult and neonatal mice.
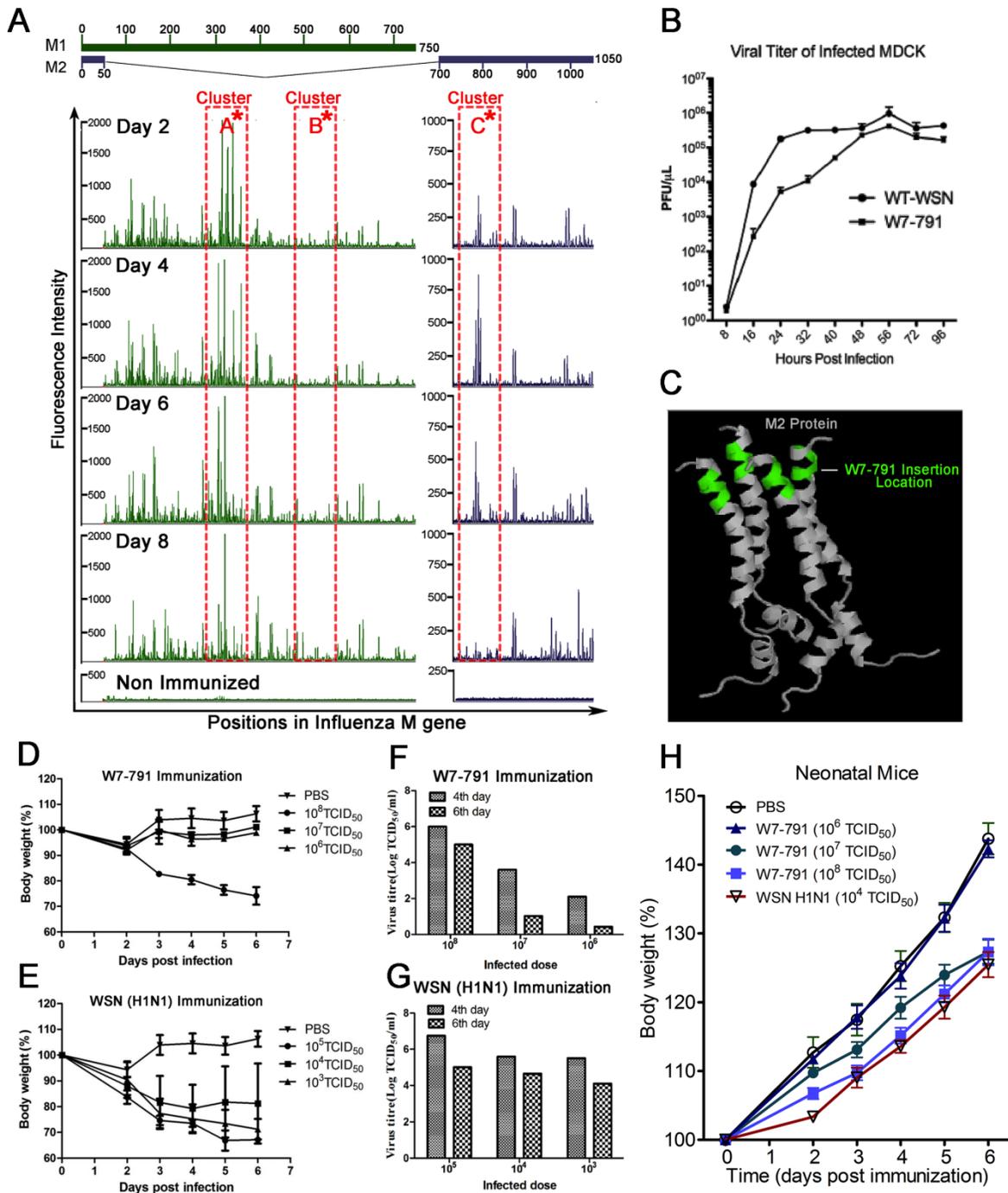
Figure 1. Generation of LAV W7-791 using genome-wide mutagenesis and *in vivo* profiling.
(A) 6-8 week old C57BL/6 mice were infected with the M gene segment mutant population by *i.t.* injection (n=8). Lungs were harvested at the indicated dpi and homogenized for genotyping. Peaks (M1 gene in green, M2 gene in blue) indicate the amount of viral RNA with insertions at that nt position. Lung homogenates from PBS- or wt WSN-injected mice serve as negative controls. Clusters A*, B* and C* are highlighted in red, representing the three types of observed growth profiles. **(B)** MDCK cells were infected at 0.25 MOI of wt WSN and W7-791 to determine the titer of virus at various time points. **(C)** The insertion position of W7-791 was mapped onto the

known M2 crystal structure in green. **(D-E)** Mouse body weight was monitored for 7 days post-inoculation with $10^6, 10^7$ or 10^8 TCID$_{50}$ of W7-791 or wt WSN. **(F-G)** The viral titers from each inoculation were evaluated on day 4 and day 6. **(H)** The weight change of BALB/c pup mice (15 days of age) inoculated with the indicated titer of W7-791, WSN or PBS.

A single dose of W7-791 can elicit protection against lethal homologous influenza virus challenge.

We next sought to determine whether vaccination with a single dose of W7-791 could protect mice against a lethal influenza virus challenge (Figure 2A). One month after *i.t.* inoculation with W7-791, mice were given a lethal dose of 4MLD$_{50}$ of wt WSN virus (Figures 2B and 2C). While the mock-vaccinated group all lost weight and died, all W7-791-vaccinated mice maintained a normal weight and survived. Interestingly, W7-791-vaccinated mice also survived without showing signs of illness or weight loss after challenge with 4MLD$_{50}$ of PR8 virus (Figures 2D and 2E).
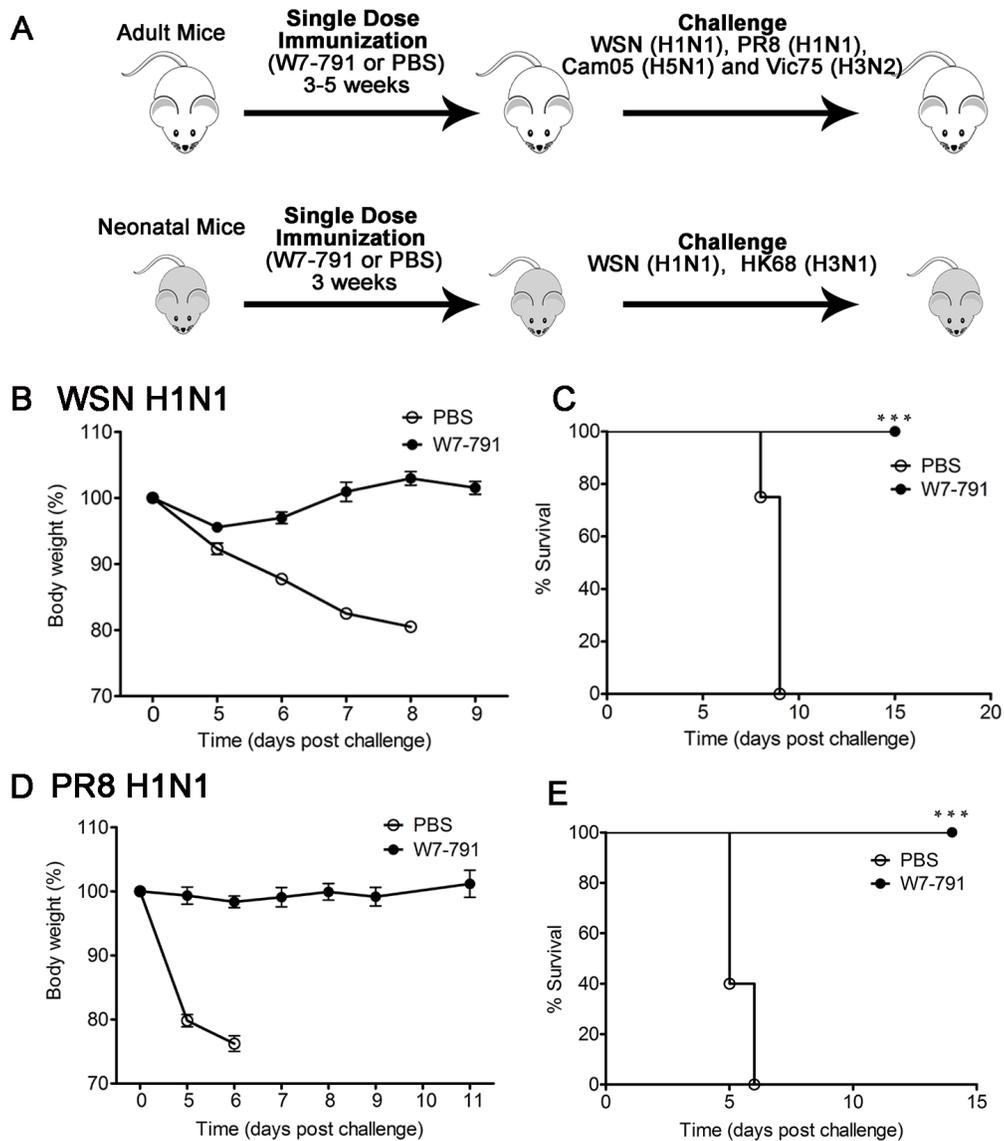
Figure 2. A single dose of W7-791 can elicit protection against lethal heterologous influenza virus challenge. (A) Schematics of strategies used for mouse immunization and challenge. **(B-C)** Mice were immunized *i.t.* with 10^5 pfu of W7-791 (n=5) or PBS (n=5). One month after vaccination, mice were challenged with $4MLD_{50}$ of WSN. Mouse weight and survival were measured at the indicated days after lethal challenge. **(D-E)** Mice were immunized *i.t.* with 10^5 pfu of W7-791 (n=5) or PBS (n=5). One month after vaccination, mice were challenged with $4MLD_{50}$ of PR8. Mouse weight and survival were measured at the indicated days after lethal challenge. *** P-value < 0.001

A single dose of W7-791 can elicit robust cross-protection against lethal heterologous influenza virus challenge.

We further explored whether W7-791 could cross-protect against heterologous highly pathogenic avian influenza (HPAI) A/Cambodia/P0322095/05 H5N1 (Cam/H5) (Figures 3A and 3B). Groups of BALB/c mice were inoculated *i.n.* with 10^6 pfu of W7-791 and challenged 3 weeks later with $2MLD_{50}$ of Cam/H5. All of the unvaccinated mice died, but the W7-791-vaccinated mice resisted the challenge without exhibiting significant weight loss. We extended our study to another phylogenetic influenza group, A/Victoria/3/75 H3N2 (Vic/H3) (Figures 3C and 3D). Mice were vaccinated with 10^5 pfu of W7-791, and then lethally challenged with $2MLD_{50}$ after 4 weeks. Interestingly, W7-791-vaccinated mice only lost about 10% of their initial weight at 3-5 dpi with Vic/H3 before fully recovering, whereas the mock-vaccinated group succumbed to the H3N2 infection. Furthermore, we challenged W7-791-immunized mice against an escalated lethal dose of $5MLD_{50}$ of WSN or HK68/H3 (Figure S4B). As before, all of the mice that were vaccinated with W7-791 survived the challenge, whereas PBS-immunized groups succumbed. Lastly, we tested whether W7-791 could cross-protect neonatal mice against lethal homologous and heterologous infections (Figures 3E and 3F). 15-day-old BALB/c mice were immunized *i.n*. with 10^6 $TCID_{50}$ of W7-791 or PBS and then challenged with a lethal dose (10^5 or 10^6 $TCID_{50}$/mice) of WSN, or a lethal dose (10^6 or 10^7 $TCID_{50}$/mice) of A/Hong Kong/68 H3N1 (HK68/H3). Again, all vaccinated mice cleared the virus and survived the infection. Lastly, we compared our W7-791 strain with a quadrivalent LAIV used over the 2015-2016 season, FluMist®. That LAIV contains attenuated viruses carrying antigens of two influenza B viruses, a H3N2 virus (Switzerland/9715293/2013), and a H1N1 strain (California/7/2009 pandemic virus). *In vivo* immunization and challenge with the same $TCID_{50}$ showed that W7-791 confers greater protection to mice against HK68/H3 (Figure S4C). All together, these results illustrate the capacity of our minimally modified mutant influenza strain to generate heterosubtypic protection against lethal virus challenge with a single immunization.
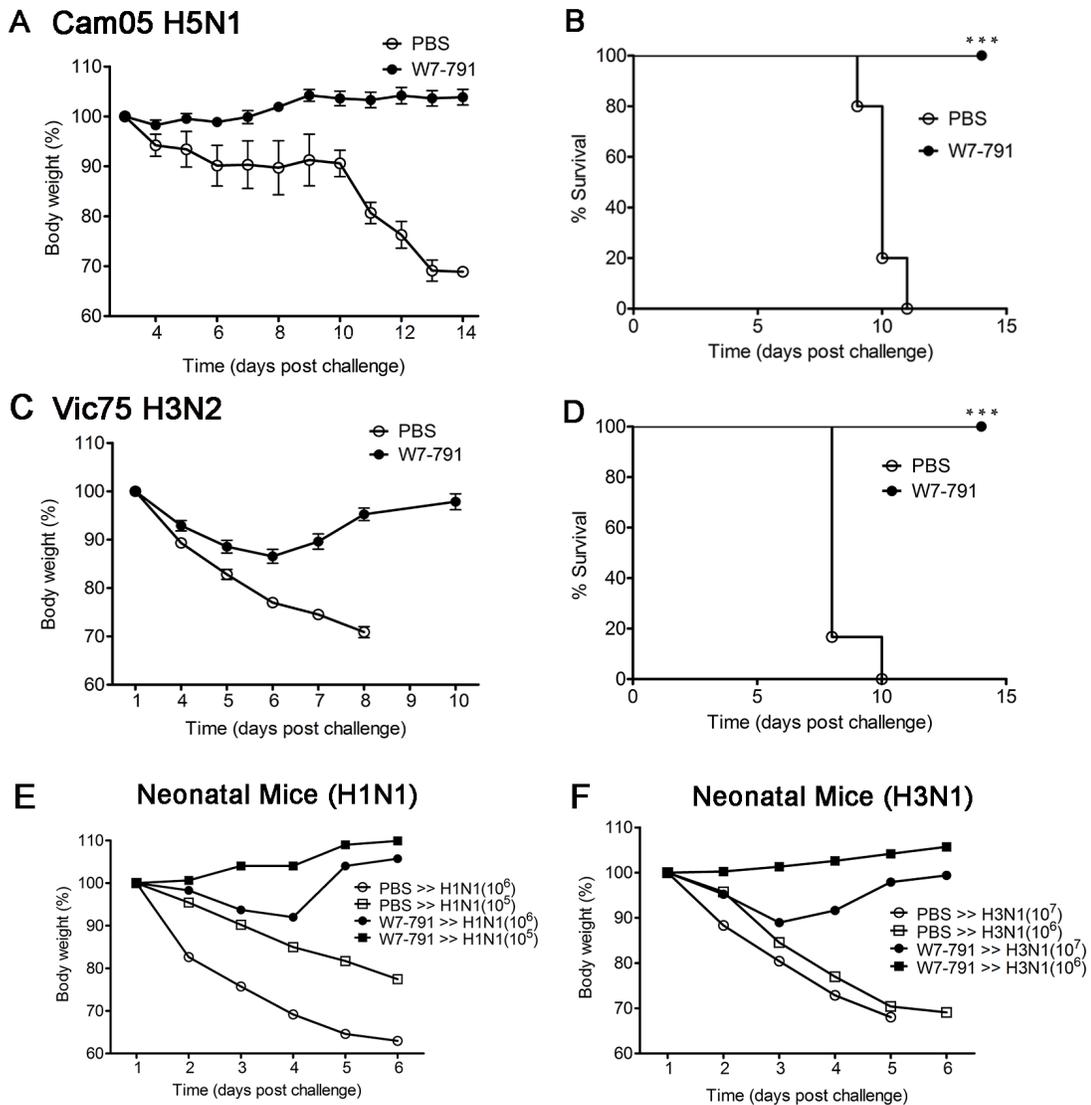
Figure 3. A single dose of W7-791 can elicit robust cross-protection against lethal heterologous influenza virus challenge. (A-B) Mice were immunized *i.n.* with 10^6 pfu of W7-791 (n=6) or PBS (n=5). 3 weeks after vaccination, mice were challenged with $2MLD_{50}$ of Cam/H5. Mouse weight and survival were measured at the indicated days after lethal challenge. **(C-D)** Mice were immunized *i.n.* with 10^5 pfu of W7-791 (n=9) or PBS (n=6). One month after vaccination, mice were challenged with $2MLD_{50}$ of Vic/H3. Mouse weight and survival were measured at the indicated days after lethal challenge. **(E-F)** Weight change of neonatal mice lethally challenged with WSN (10^5 or 10^6 $TCID_{50}$) or HK68/H3 (10^6 or 10^7 $TCID_{50}$) 3 weeks after challenge. *** P-value < 0.001

W7-791 activates both humoral and cell-mediated immune responses.

Live attenuated vaccines are thought to confer broad cross-protection against heterologous strains through activation of both humoral and cell-mediated immune responses. To determine the mechanisms responsible for the protective effects of W7-791, we immunized mice *i.n.* with W7-791, then challenged with WSN or PR8 viruses. We found that W7-791-vaccinated mice had significantly improved viral clearance against both viruses (Figure 4A). Surprisingly, W7-791-vaccinated mouse serum only showed HAI against WSN and not against heterologous strains PR8, HK68/H3, A/Wisconsin/2005 H3N2 (Wis/H3) or Cam/H5 pseudotype virus (Figures 4B and C). We further tested the serum using a micro-neutralization assay against WSN, PR8 and HK68/H3. Mice vaccinated with W7-791 showed a high neutralization titer against WSN and a lower level against PR8 and HK68/H3 (Figure 4D). These findings led us to believe that antibodies may not be the only source of protection mediated by W7-791.

To differentiate the cross-protective effect of cell-mediated and humoral responses, we immunized mice *i.n.* with 10^6 pfu of W7-791 or PBS, and collected serum and total T cells from spleens and lymph nodes one month later. We transferred these separately to naïve mice and challenged them *i.n.* with a lethal dose of WSN or HK68/H3. The directly vaccinated (not transferred) group was used as a positive control and showed 100% protection. Adoptive transfer of serum from W7-791-immunized mice gave full protection against WSN, but not against HK68/H3 (Figure 4E). However, adoptive transfer of T cells from W7-791-immunized mice was able to partially or fully protect naïve mice from HK68/H3 or WSN, respectively (Figure 4F). This demonstrates that W7-791 vaccination can generate a T cell-mediated immune response with a heterosubtypic protective ability. Furthermore, we evaluated CD8 T cell epitopes in the W7-791-vaccinated mice (data not shown). We found that a large proportion of lung CD8+ T cells in W7-791-immunized mice were specific to H-2^{Db}-ASNENMETM (NP) 7 days after a H3N1 challenge. These data, combined with our adoptive T cell transfer experiment (shown in Figures 3F and H),

has led us to believe that W7-791 immunization elicits cross-protective CD8+ T cell responses in our mouse model.
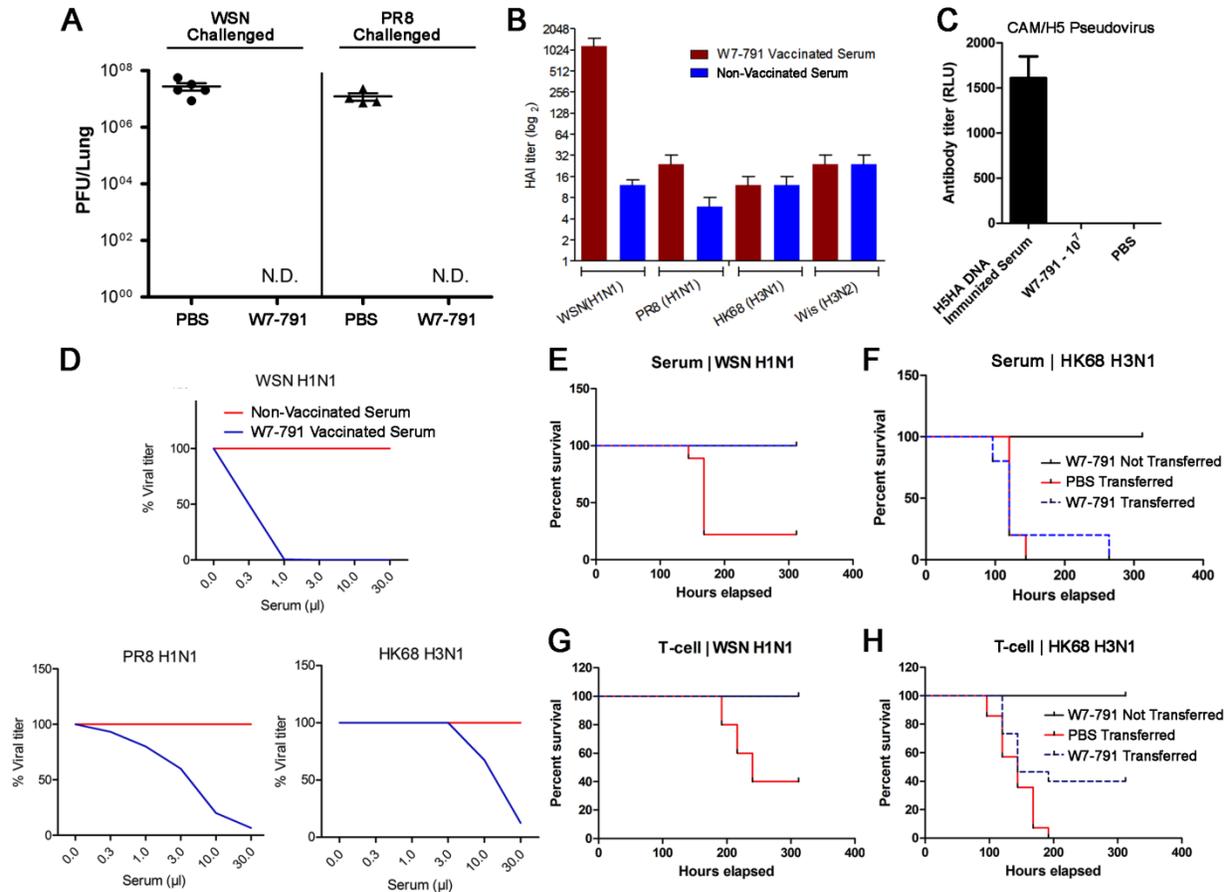
Figure 4. W7-791 activates both humoral and cell-mediated immune responses. (A) Viral titers in lung homogenates of vaccinated and unvaccinated mice, challenged with lethal WSN or PR8, were quantified by plaque assay at the time of death or euthanasia. **(B)** HAI assay of vaccinated and unvaccinated mouse serum against WSN, PR8, HK68/H3 and A/Wisconsin/05 H3N2. **(C)** The level of antibody in vaccinated and unvaccinated mouse serum was measured using Cam/H5 pseudovirus neutralization. Measurements are in relative light units (RLU). **(D)** Microneutralization assay of W7-791 vaccinated or unvaccinated mouse serum against WSN, PR8 and HK68/H3. **(E-F)** Survival of mice receiving serum from W7-791 (n=14) or PBS-vaccinated mice (n=10). Mice were challenged 24 hours later with a lethal dose of WSN or HK68/H3. **(G-H)** Survival curve of mice receiving T-cells from W7-791 (n=20) or PBS immunized mice (n=19), subsequently challenged with a lethal dose of WSN or HK68/H3. W7-791-vaccinated but not transferred mice were used as a control (n=10).

A single dose of W7-791 elicits heterologous protection in the ferret model.

To further explore the efficacy of W7-791 as a LAIV strain, we selected the ferret model, which resembles the pathophysiology of human influenza infection more closely than the mouse (Maher and DeStefano, 2004; Matsuoka et al., 2009). To evaluate the clinical response, we inoculated ferrets with 10^6, 10^7 and 10^8 TCID$_{50}$ of W7-791. We observed no significant changes in temperature (Figure 5A) or clinical symptoms (Figure 5B) after inoculation with 10^8 TCID$_{50}$ of W7-791. W7-791-vaccinated ferrets showed a significant increase in serum antibody titer (Figure 5C), but HAI assay confirmed that these antibodies bind to only the HA from WSN and not HK68/H3 (Figures 5D) or H5N1 (Figure S5A). To examine protection against heterologous viruses, we challenged the ferrets 4 weeks post-immunization with either 10^6 TCID$_{50}$ of WSN, 10^6 TCID$_{50}$ of HK68/H3 or PBS. Ferrets that had been immunized with 10^3 and $10^{4.7}$ TCID$_{50}$ of W7-791 showed significant reduction in shedding of the challenge virus within 2 dpi, compared to the unvaccinated group (control WSN or HK68/H3 respectively) (Figures 5E-F and S5B). The vaccinated ferrets displayed an undetectable WSN or HK68/H3 viral load at 2 or 4 dpi, respectively. The vaccinated ferrets also showed improved clinical symptoms compared to non-vaccinated ferrets (Figures 5G-H). These data suggest that a single dose of W7-791 is safe in ferrets and can also elicit heterologous subtypic protection.
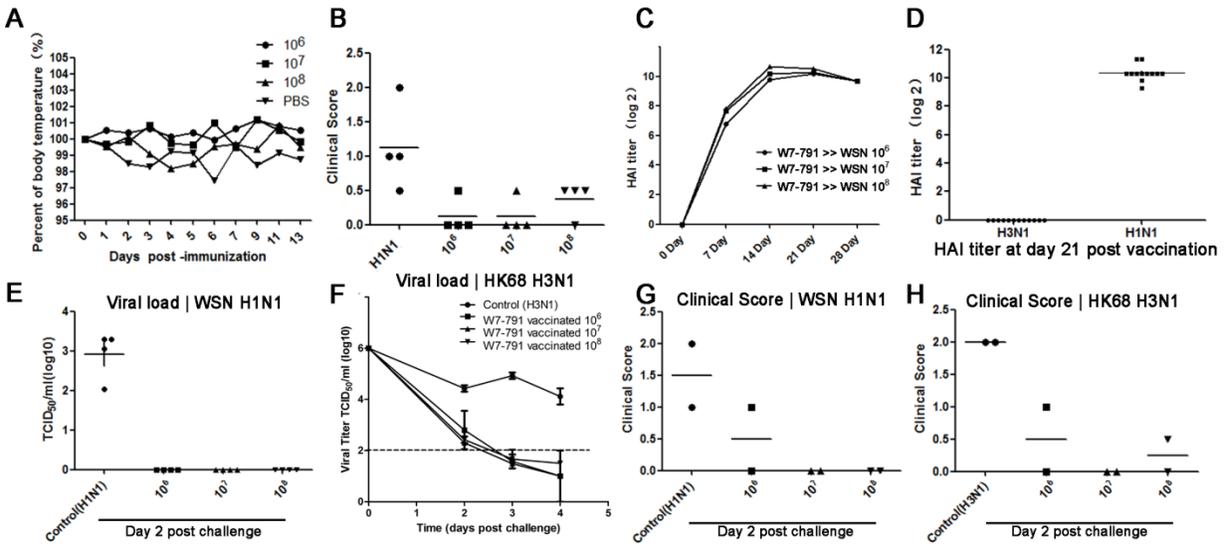
Figure 5. A single dose of W7-791 elicits heterologous protection in the ferret model. (A) Temperature curve of ferrets inoculated *i.n.* with 10^6, 10^7 or 10^8 TCID$_{50}$ of W7-791 or PBS. **(B)** Clinical scores of the ferrets 2 and 3 days post-inoculation with indicated doses of W7-791 or 10^6 TCID$_{50}$ of WSN. Clinical signs were evaluated based on Reuman *et al (20).* **(C)** HAI assay showing the increase of anti-W7-791 HA antibody titer from W7-791 inoculated ferrets after vaccination. **(D)** HAI assay showing the titer of anti-H1HA or H3HA antibody 21 days post-inoculation. **(E-F)** The viral titers **(G-H)** and the clinical scores of the immunized and non-immunized ferrets challenged with 10^6 TCID$_{50}$ of WSN or HK68/H3.

3.2.c DISCUSSION

We have taken the approach of combining genome-wide mutagenesis and *in vivo* growth profile screening for the rapid and high-throughput development of LAIVs. As proof of principle, we used this method to uncover a promising influenza LAIV, W7-791, that is over 100-fold more attenuated than the parent H1N1 virus, and is even tolerated in neonatal mice with no observable lung pathology. This strain could also confer complete cross-protection against heterologous H3 and H5 subtype influenza strains in mice. We further demonstrated that the insertion could be broadly applicable to attenuate other clades of influenza virus. The knock-out or modified influenza M2 has been previously described as a potential LAIV candidate (Hatta et al., 2011; Watanabe et al., 2009). This present study explored the untapped potential of the M protein in generating LAIVs. The quadrivalent LAIV (FluMist®) that was administered in the U.S. over the previous three seasons was shown to be ineffective compared to the inactivated vaccine. Our own comparison between the 2015-2016 FluMist formulation and W7-791 showed that our mutant H1 strain could confer greater protection in mice against challenge with a H3 virus than FluMist, which itself contains a H3 strain component. One possible explanation is that the primary method of protection in this case was not via H3-specific epitopes that were otherwise present in the quadrivalent LAIV, but rather through conserved epitopes still present in our H1 mutant strain. It is also plausible that W7-791, which is derived from a mouse-adapted strain, stimulated the immune system more efficiently through active replication in mice, and whether it can out-perform FluMist in primates remains to be investigated. Although the reason for the failure of the FluMist over the last three seasons is under investigation, it was observed that this failure came at a time when the formulation was changed to having four viral components. While the inactive quadrivalent vaccine formulation is still effective, it is possible that including too many live influenza strains in the vaccine causes growth competition and reduced infectivity. Therefore, our strain may simply have been more effective because it was a single component vaccine. The failure of FluMist is unfortunate, as live attenuated vaccines induce cell-mediated responses that are thought to target the more conserved parts of the virus, giving them the greatest potential for

inducing broad protection against a multitude of strains. This would not only bypass the need to adjust our vaccines each year to account for antigenic drift variants, but could also protect us from highly pathogenic avian strains and novel reassortment viruses that could become the next influenza pandemic. W7-791 has shown great promise in its ability to confer protection against lethal heterologous influenza H1, H3 and H5 strains in mice (including neonates) and ferrets. We demonstrated that even a single-dose of W7-791 in mice could induce protection, without the need for prime-boost strategies or adjuvants. However, we acknowledge that the current extent of our LAIV characterization is limited. Future studies could include transferring the insertion to a master donor strain used for LAIVs, and expanding the range of the challenge doses to further evaluate the potential of this candidate vaccine. It also remains to be seen whether W7-791 will be successful in animals that have been previously exposed to influenza virus. In short, the development of a better influenza vaccine will require much more investigation, and the present work is only an initial step.

Our study also advances our understanding of the roles of antibody and T cell responses in influenza infection. Furthermore, we hypothesize that our strategy of screening for a growth phenotype that permits limited replication in the host and can elicit both cell-mediated and humoral immune responses without causing illness, may be applicable to finding attenuated vaccines for other viruses. It would also be interesting to compare how the innate and adaptive immune systems communicate in response to influenza infection, a LAIV, or an inactivated vaccine.

This work presents a LAIV development strategy that can rapidly generate and screen entire libraries of viral clones for attenuation. We associated a particular viral growth profile in mice with the ability to elicit both cell-mediated and humoral immune responses without causing significant disease. From a cluster of viruses showing this ideal level of attenuation, we identified a promising LAIV candidate. This approach could potentially be implemented to discover attenuated mutants for other RNA and DNA viruses. It would require no working knowledge of

specific genes or their functions in the virus, but could be used to establish a comprehensive profiling of the entire genome *in vitro* and *in vivo*. This system would also be ideal for expediting the design of live attenuated vaccines against less understood or emergent viruses.

MATERIAL AND METHODS

Cell culture

HEK293T cells were cultured in Dulbecco's modified Eagle's medium supplemented with 5% heat-inactivated fetal bovine serum (FBS). Madin-Darby canine kidney (MDCK) cells were maintained in Dulbecco's minimal essential medium (DMEM) containing 5% fetal bovine serum, penicillin/streptomycin (100 units/ml and 50 μg/ml, respectively), and 1 mM sodium pyruvate at 37°C with 5% CO_2.

Generation of M gene segment mutant plasmid library and functional profiling

To create the mutant plasmid library of the M gene segment of influenza A virus A/WSN/1933, a 15-nt sequence (5'-NNNNNTGCGGCCGCA-3'; N=duplicated 5 nucleotides from target DNA) was randomly inserted by Mu-transposon mediated mutagenesis (MGS kit, Finnzymes) according to the manufacturer's instructions. The M gene mixed mutant pool was transformed into *E. coli* DH10B by electroporation at 2.0kV, 200Ω, 25μF (ElectroMax [TM] DH10B, Invitrogen). The mutant M gene plasmid and 7 remaining wt plasmids were transfected concomitantly into HEK293T cells for virus generation. Three days after transfection, the supernatant was collected and transferred to MDCK cells for propagation. Virus was collected after 48 hrs, then either stored or used for further propagation for up to 4 passages. RNA was isolated with the TRIzol reagent (Invitrogen) after each generation. Reverse transcription-PCR was carried out with the iScript[TM] cDNA Synthesis kit (Bio-Rad) to create cDNA. Three gene-specific forward primers approximately 400bp apart in the M-gene segment (5'-AGCAAAAGCAGGTAGATATT-3', 5'-GGGGCCAAAGAAATAGCACT-3', 5'-TCCTAGCTCCAGTGCTGGTC-3') and a Vic-labeled insertion-specific mini-primer (5'-TGCGGCCGCA -3') were used to amplify fragments containing the 15nt insert using KOD Hot-Start polymerase (Novagen). The PCR conditions were set to 95°C for 10 min (1 cycle), 95°C for 45sec, 52°C for 30sec and 72°C for 90sec (30 cycles), and 72°C for 10 min (1 cycle).   The fluorescent-labeled PCR products were analyzed in duplicate with a Liz-500 size standard (Applied Biosystems) using a 96-capillary genotyper (3730xl DNA Analyzer,

Applied Biosystems) at the UCLA GenoSeq Core facility. Sequencing data was analyzed for clarity using ABI software, with the following criteria: (1) All data passed the standard default detection level; (2) The first 70bp were removed due to non-specific background noise; (3) All data were aligned to the nearest base pair in the influenza A WSN matrix gene; (4) All genotyping experimental data was normalized with wt WSN-infected cells, non-transfected cells and a different gene library as controls. This eliminated non-specific data from the PCR, primers and the DNA Analyzer. For infection *in vivo,* the mutant virus pool was titered, concentrated by ultra-centrifugation and re-titered, and used for mouse injection. Two days post-infection, the lungs were harvested, homogenized and resuspended in TRIzol for RNA isolation, followed by the same procedures as described above. PBS or WSN-infected mice served as controls.

Virus stains

We used the influenza A/WSN/1933 reverse genetics system to generate seasonal A/H1N1 virus (Hoffmann et al., 2000). This strain is a mouse-adapted influenza virus and has been used as the parental strain to generate potential LAIVs using transposon mutagenesis. The 8 plasmids containing the cDNA of A/WSN/33 (gift from Dr. Yuying Liang at Emory University) were transfected into HEK293T cells using TransIT LT-1 (Panvera, Madison, WI) by the manufacture's protocol. The virus was serially passaged three times in MDCK cells to a final titer of $10^{7.4}$ PFU/ml. Influenza virus A/Puerto Rico/8/1934 (seasonal A/H1N1 virus) was a gift from Dr. Yuying Liang. The virus was serially passaged three times in MDCK cells to a titer of $10^{7.5}$ PFU/ml. The 50% mouse lethal dose (MLD_{50}) of both strains determined in C57BL/6 mice.

Influenza virus A/Victoria/3/75 (seasonal A/H3N2 virus), A/Wisconsin/65/05 (seasonal A/H3N2 virus), and A/Hongkong/68 (seasonal A/H3N1 virus) were gifts from Dr. Ioanna Skountzou at Emory University. These viruses were amplified using MDCK cells for 2-3 passages to a final titer of $10^{5.5}$ PFU/ml, $10^{5.4}$ PFU/ml and 10^7 PFU/ml, respectively. The MLD_{50} was determined in C57BL/6 and BALB/c mice.

Influenza virus A/Cambodia/P0322095/05 (Highly Pathogenic Avian Influenza H5N1 virus) was originally isolated from human patients at the Pasteur Institute in Cambodia (Buchy et al., 2007). Virus was propagated in MDCK cells and virus-containing supernatants were pooled, clarified by centrifugation and stored at -80°C. The 50% tissue culture infection dose ($TCID_{50}$) and the MLD_{50} of the viruses were determined in MDCK cells and in BALB/c mice, respectively, and were calculated as described previously (Ding et al., 2011).

Virus Titrations

The concentration of infectious viruses was determined by plaque assay and end-point titrations. Plaque assays were performed in MDCK cells and calculated as plaque forming units per µL (pfu/ L) of supernatant. The viral samples were serially diluted in dilution buffer (PBS with 10% BSA, $CaCl_2$, 1% DEAE-dextran, and $MgCl_2$). Diluents were added to a monolayer of MDCK cells in 6-well plates for 1 hour at 37°C, and then covered with growth medium containing 1% low-melting agarose and TPCK-treated trypsin (0.7 µg/ml). Infected cells were stained after 48 hours (1% crystal violet, 20% ethanol, in PBS) to visualize the plaques. Virus titrations were performed by end-point titration in MDCK cells. MDCK cells were inoculated with tenfold serial dilutions of the virus, then washed with PBS once 1 hr after inoculation, and cultured in DMEM for 48 hrs to visualize cell viability. The viral titer was determined by luminescence assay or by plaque assay. To measure the growth of individual mutants *in vitro* (Figure S2B), an influenza virus-responsive *Gaussia* luciferase (gLuc) reporter system was used. Briefly, the gLuc coding region was inserted in the reverse-sense orientation between a human RNA polymerase I promoter and a murine RNA polymerase I terminator. The gLuc coding sequence was flanked by the UTRs from the PA segment of influenza virus A/WSN/33 strain so that gLuc expression is dependent on influenza virus infection. The gLuc reporter was transfected into HEK293Ts for 24hours before the supernatants containing mutant or wt influenza viruses were added. Upon active infection, gLuc is released into the supernatant and can be quantified with *Renilla* luciferase substrate (Promega).

Animals

Adult mice

Female C57BL/6 mice, 6-8 weeks old, were purchased from the Jackson Laboratory. All animals were housed in pathogen–free conditions within the UCLA animal facilities.

Neonatal mice

15day-old BALB/c mice (Vital River Beijing, China) weighing 6-9 grams were inoculated intranasally (*i.n.*) with PBS, 10^4 TCID50 of WSN virus, or dilutions of W7-791. For the dose-dependent experiment, mice were inoculated *i.n.* with 10^6, 10^7 and 10^8 TCID$_{50}$ of W7-791. 16 days post-treatment, mice were challenged *i.n.* with a lethal dose (10^5 or 10^6 TCID$_{50}$/mouse) of WSN or (10^6 or 10^7 TCID$_{50}$/mouse) A/Hong Kong/68 H3N1 (HK68/H3) in a 30µl volume. Randomly selected mice from each group were sacrificed for pathological examinations of the lung at 4 and 6 dpi. Then the lungs were homogenized to measure viral titer using end-point-dilution assays.

Ferrets

Healthy young adult outbred female ferrets (Mustela putorius furo; between 4-5 months of age) were purchased from a commercial breeder (Wuxi, China) and confirmed to be seronegative by hemagglutination inhibition (HAI) assay to A/WSN/1933 (H1N1), A/Victoria/3/75 (H3N2), HK68 (H3N1) and W7-791(H1N1). A minimum of three independently housed ferrets were inoculated *i.n.* with 0.5 ml (0.25 ml per nostril) of 10^6, 10^7, or 10^8 TCID$_{50}$ of W7-791 or PBS. Anesthesia was performed on the quadriceps muscles of the left hind leg with a total volume of 0.02 ml Lumianning (Hua Mu Animal Care, Jiling). Serum samples were collected at day 0, 7, 14, 21, and 28 post-immunization for HAI studies. Nasal washes were collected 0-7 days after immunization. 4 weeks after immunization, the ferrets were challenged *i.n.* with 10^6 TCID$_{50}$ of WSN (H1N1) or HK68 (H3N1). Weights and temperatures were monitored daily for 7 days after inoculation. Nasal washes were collected 0-7 days after the challenge. Clinical signs were evaluated 3 days prior to vaccination, then 9, 11, 13 and 15 dpi, and 2 days prior to challenge and 1-7 dpi. The clinical

signs were scored as previously described (Reuman et al., 1989). All animal studies were performed according to the guidelines of the UCLA Animal Research Committee.

Mouse immunization and challenge

Female C57BL/6 and BALB/c mice were randomly divided into groups of 5 or 6 mice. Groups were inoculated *i.n.* or *intratracheally (i.t.)* with either PBS or W7-791 in a volume of 50µl. Intratracheal injection was performed by anesthetizing mice *i.p.* with a ketamine/xylazine mixture, then surgically exposing the trachea for direct injection of 30µl of solution with a sterile 27G needle (Shahangian et al., 2009). 4 weeks after immunization, all mice were challenged *i.n.* or *i.t* with an influenza strain in a 50µl volume: A/WSN/1933 (H1N1) at $4MLD_{50}$, A/Puerto Rico/8/1934 (H1N1) at $4MLD_{50}$, A/Cambodia/P0322095/05 (HPAI-H5N1) at $2MLD_{50}$, or A/Victoria/3/75 (H3N2) at $2MLD_{50}$. Mice were monitored and recorded daily for signs of illness, such as lethargy, ruffled hair and weight loss. When mice lost 30% or more of their original weight, they were euthanized and counted as dead. For the adoptive transfer experiment, female C57BL/6 mice were randomly divided into two sets of vaccinated or unvaccinated groups. Unvaccinated mice were sham immunized, whereas the vaccinated group received a single-dose of W7-791 at 10^6 pfu/mouse. One set from each group was used to harvest cells for the transfer experiment 4 weeks post-vaccination, while the other set was used as a vaccinated but not transferred control. Total CD4+ and CD8+ T cells were isolated from the spleens of the vaccinated and the unvaccinated mice using the Mouse Pan T Cell Isolation Kit and MS columns (Miltenyi Biotec). On the same day, the cells from the same group were pooled, and ~$10^{6.3}$ T cells/mouse were injected via the retro-orbital route to a new set of naïve female C57BL/6 mice. Likewise, sera was isolated from either the vaccinated or unvaccinated groups and matching groups were pooled, then 100µl/mouse of serum was administered retro-orbitally to a new set of naïve female C57BL/6 mice. The mice in all groups were challenged *i.n.* at 24 hours post-adoptive transfer with $2MLD_{50}$ of WSN or $2MLD_{50}$ of HK68/H3.

In vivo challenge using HPAI virus H5N1

All animal protocols were approved by the Institutional Animal Care and Use Committee at the Pasteur Institute of Cambodia. Female BALB/c mice (*Mus musculus*) at the age of 6 to 8 weeks were purchased from Charles River Laboratories (L'Arbresle, France) and housed in microisolator cages ventilated under negative pressure with HEPA-filtered air and a 12/12-hour light/dark cycle. Virus challenge studies were conducted in BSL3 facilities at the Pasteur Institute of Cambodia. Before each inoculation or euthanasia procedure, the mice were anesthetized by intraperitoneal (*i.p.*) injection of pentobarbital sodium (75 mg/kg; Sigma).

Ethical Statement

All animal experiments were carried out at biosafety level 3 (BSL3) containment facilities complying with the Ethics Committee regulations of the Institut Pasteur, Paris, France, in accordance with EC directive 86/609/CEE and were approved by the Animal Ethics Committee of the Institut Pasteur in Cambodia (permit number VD100820). Before each inoculation or euthanasia procedure, the mice were anesthetized by *i.p.* injection of pentobarbital sodium, and all efforts were made to minimize suffering.

Lung homogenization

After animals were sacrificed, lungs were perfused by injecting 1 ml of PBS containing 5 mM EDTA into the right ventricle. Whole lungs were removed and the lymph nodes were dissected away. The lungs were homogenized with 1 ml of PBS containing a proteinase inhibitor cocktail (Roche Applied Science), and virus titers in lungs were evaluated by plaque assay. After homogenates were centrifuged at 10,000 x g for 10 min, the supernatant was collected for genotyping.

Sequence comparisons

Influenza A Matrix 1 and Matrix 2 protein sequences from approx. 300 previously reported strains from 1918-2014 were compared and aligned using the NCBI (Influenza database, http://www.ncbi.nlm.nih.gov/genomes/FLU/aboutdatabase.html).

Structure analysis

Conserved and viable mutations in the M-gene were mapped onto the crystal structure of the monomeric M1 gene (PDB accession ID, 2Z16) and the tetrameric M2 gene (PDB accession ID, 2L0J), which were obtained from the protein data bank (PDB). The structure labeling was performed using PyMOL v.1.0.

***In vitro* assays**

Cell viability assay

Cell viability was measured by CytoTox 96 Non-Radioactive Cytotoxicity Assay (Promega) according to the manufacturer's instructions.

Hemagglutination inhibition (HAI) assay

Viruses A/WSN/1933, A/Puerto Rico/8/1934, A/Wisconsin/65/05 and A/Hong Kong/68 were diluted to 4 HA units and incubated with an equal volume of serially diluted sera for 30 min at room temperature. An equal volume of 1% chicken red blood cells was added to the wells and incubation continued on a gently rocking plate for 30 min at room temperature. Button formation was scored as evidence of hemagglutination inhibition. Assays were performed in triplicate.

Microneutralization assay

MDCK cells (5×10^5 cells per well) were seeded onto a 12-well culture plate in complete DMEM overnight. To test the neutralization activity of immune sera, serial 3-fold dilutions of sera were incubated with $10^{6.5}$ PFU/ml, $10^{4.4}$ PFU/ml, $10^{4.2}$ PFU/ml of viruses A/WSN/1933, A/Hongkong/68

and A/Puerto Rico/8/1934 at the final volume of 100 µl at room temperature for 1 hour. After the

incubation, the mixture was added onto a monolayer of MDCK cells and was incubated for 1 hour

at 37°C and then covered with growth medium containing 1% low-melting-point agarose and

TPCK-treated trypsin (0.7 µg/ml). Infected cells were stained after 48 hours (1% crystal violet, 20%

ethanol, in PBS) to visualize the plaques. Assays were performed in triplicate.

Pseudovirus neutralization assay

H5N1 pseudotype virus expressing the H5HA derived from A/Cambodia/P0322095/05

(Accession# of HA: ADM95463), the N1NA (Accession# of NA: AY555151) derived from

A/Thailand/1(KAN-1)/2004 and a luciferase reporter gene were used in this experiment. The

ferret sera were diluted in 2-fold serial dilutions from 1/20 to 1/1280 and the mouse sera were

diluted from 1/10 to 1/1280. Sera from mice immunized by injection of H5HA DNA (Accession#

of HA: AAS65615) from A/Thailand/1(KAN-1)/2004 was used as a positive control. IC_{50} values

were defined as the dilution of a given immune serum that resulted in 50% reduction of RLA.

Lung homogenization

After animals were sacrificed, lungs were perfused by injecting 1 ml of PBS containing 5 mM

EDTA into the right ventricle. Whole lungs were removed and the lymph nodes were dissected

away. The lungs were homogenized with 1 ml of PBS containing a proteinase inhibitor cocktail

(Roche Applied Science), and virus titers (pfu) in lungs were evaluated by plaque assay. After

homogenates were centrifuged at 10,000 x g for 10 min, the supernatant was collected for

genotyping.

Sequence comparisons

Influenza A Matrix 1 and Matrix 2 protein sequences from approx. 300 previously reported strains

from 1918-2014 were compared and aligned using NCBI (Influenza database,

http://www.ncbi.nlm.nih.gov/genomes/FLU/aboutdatabase.html).

Structure analysis

Conserved and viable mutations in the M-gene were mapped to crystal structure of the monomeric M1 gene (PDB accession ID, 2Z16) and the tetrameric M2 gene (PDB accession ID, 2L0J), which were obtained from protein data bank (PDB). The structure labeling was performed using PyMOL v.1.0.

REFERENCE:

1.   WHO (2003) Influenza: Fact sheets.

2.   Belshe RB, *et al.* (2007) Live attenuated versus inactivated influenza vaccine in infants and young children. *N Engl J Med* 356(7):685-696.

3.   Lamb RA, Lai CJ, & Choppin PW (1981) Sequences of mRNAs derived from genome RNA segment 7 of influenza virus: colinear and interrupted mRNAs code for overlapping proteins. *Proc Natl Acad Sci U S A* 78(7):4170-4174.

4.   Krammer F & Palese P (2015) Advances in the development of influenza virus vaccines. *Nature reviews. Drug discovery* 14(3):167-182.

5.   Tompkins SM, *et al.* (2007) Matrix protein 2 vaccination and protection against influenza viruses, including subtype H5N1. *Emerg Infect Dis* 13(3):426-435.

6.   Ilyinskii PO, *et al.* (2008) Inhibition of influenza M2-induced cell death alleviates its negative contribution to vaccination efficiency. *PLoS One* 3(1):e1417.

7.   Schnell JR & Chou JJ (2008) Structure and mechanism of the M2 proton channel of influenza A virus. *Nature* 451(7178):591-595.

8.   Neirynck S, *et al.* (1999) A universal influenza A vaccine based on the extracellular domain of the M2 protein. *Nature medicine* 5(10):1157-1163.

9.   De Filette M, *et al.* (2008) An influenza A vaccine based on tetrameric ectodomain of matrix protein 2. *Journal of Biological Chemistry* 283(17):11382-11387.

10.   Schotsaert M, De Filette M, Fiers W, & Saelens X (2009) Universal M2 ectodomain-based influenza A vaccines: preclinical and clinical developments. *Expert review of vaccines* 8(4):499-508.

11.   Wang L, *et al.* (2014) Nanoclusters self-assembled from conformation-stabilized influenza M2e as broadly cross-protective influenza vaccines. *Nanomedicine : nanotechnology, biology, and medicine* 10(2):473-482.

12.   Huleatt JW, *et al.* (2008) Potent immunogenicity and efficacy of a universal influenza vaccine candidate comprising a recombinant fusion protein linking influenza M2e to the TLR5 ligand flagellin. *Vaccine* 26(2):201-214.

13.   De Filette M, *et al.* (2005) Universal influenza A vaccine: optimization of M2-based constructs. *Virology* 337(1):149-161.

14.   Arumugaswami V, *et al.* (2008) High-resolution functional profiling of hepatitis C virus genome. *PLoS Pathog* 4(10):e1000182.

15.   Maher JA & DeStefano J (2004) The ferret: an animal model to study influenza virus. *Lab animal* 33(9):50-53.

16.   Matsuoka Y, Lamirande EW, & Subbarao K (2009) The ferret model for influenza. *Current protocols in microbiology* Chapter 15:Unit 15G 12.

17. Hoffmann E, Neumann G, Kawaoka Y, Hobom G, & Webster RG (2000) A DNA transfection system for generation of influenza A virus from eight plasmids. *Proc Natl Acad Sci U S A* 97(11):6108-6113.

18. Buchy P*, et al.* (2007) Influenza A/H5N1 virus infection in humans in Cambodia. *J Clin Virol* 39(3):164-168.

19. Ding H*, et al.* (2011) Superior neutralizing antibody response and protection in mice vaccinated with heterologous DNA prime and virus like particle boost against HPAI H5N1 virus. *PLoS One* 6(1):e16563.

20. Reuman PD, Keely S, & Schiff GM (1989) Assessment of signs of influenza illness in the ferret model. *J Virol Methods* 24(1-2):27-34.

21. Shahangian A*, et al.* (2009) Type I IFNs mediate development of postinfluenza bacterial pneumonia in mice. *J Clin Invest* 119(7):1910-1920.

Fig. S1.
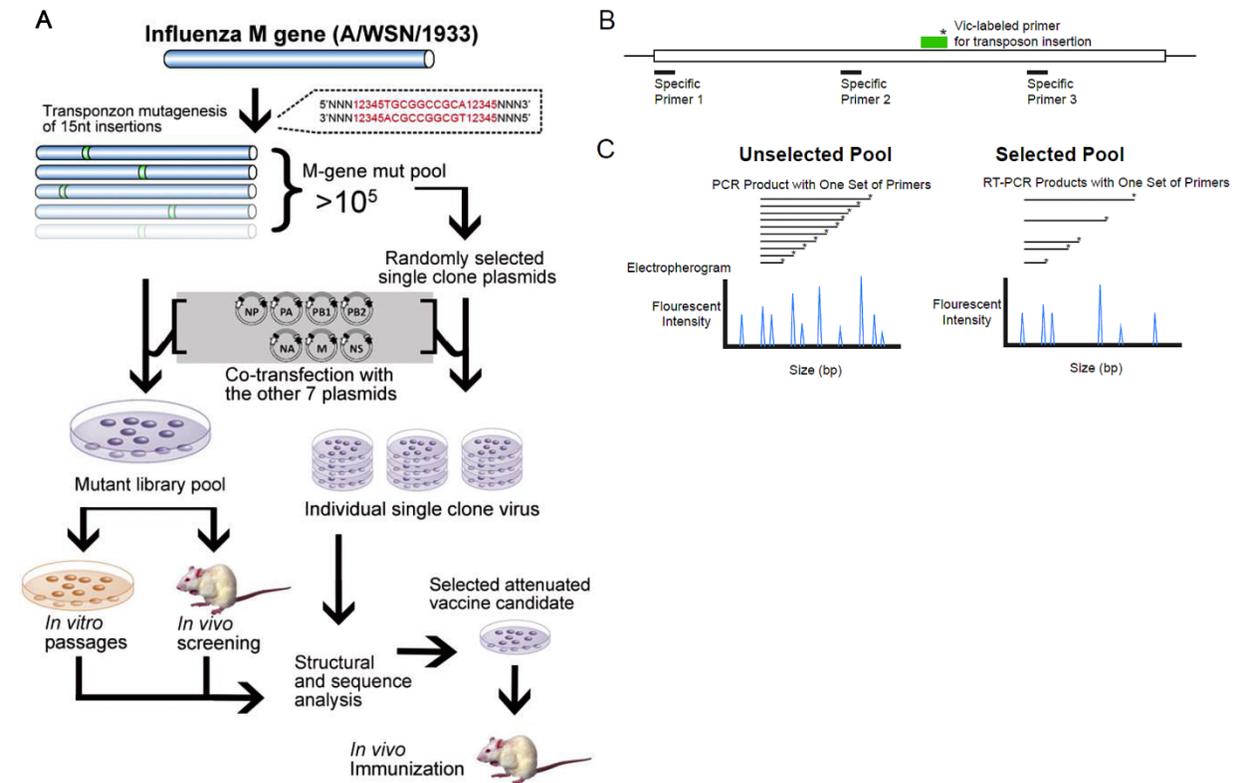
Supplementary Figure 1. Generation of the transposon-based M gene segment mutant library. (A) High-density mutations were made in the M gene segment of influenza (A/WSN/1933) through the random insertion of 15nt by transposon mutagenesis. Transfection of HEK293T cells with the mutant M plasmids and seven wt plasmids of other influenza genes was used to generate the mutant virus library. Supernatants containing the mutant viral library were used to infect MDCK cells (*in vitro*) or mice (*in vivo*). Lung and lymphoid organs from the inoculated mice were collected at various time points to recover virus. Mutations in regions that led to viral growth attenuation were identified by genotyping and further characterized *in vitro* and *in vivo*. **(B)** Genotyping by PCR amplification using gene-specific primers and a primer against the 15nt insert will show the location of all mutants in the pool. **(C)** The mutant pool library can be put through selection such as growth *in vitro* or *in vivo*. Essential and non-essential regions in the genome required for growth can be determined by comparing genotyping data from the unselected pool (total mutant library) and selected pool.

Fig. S2.
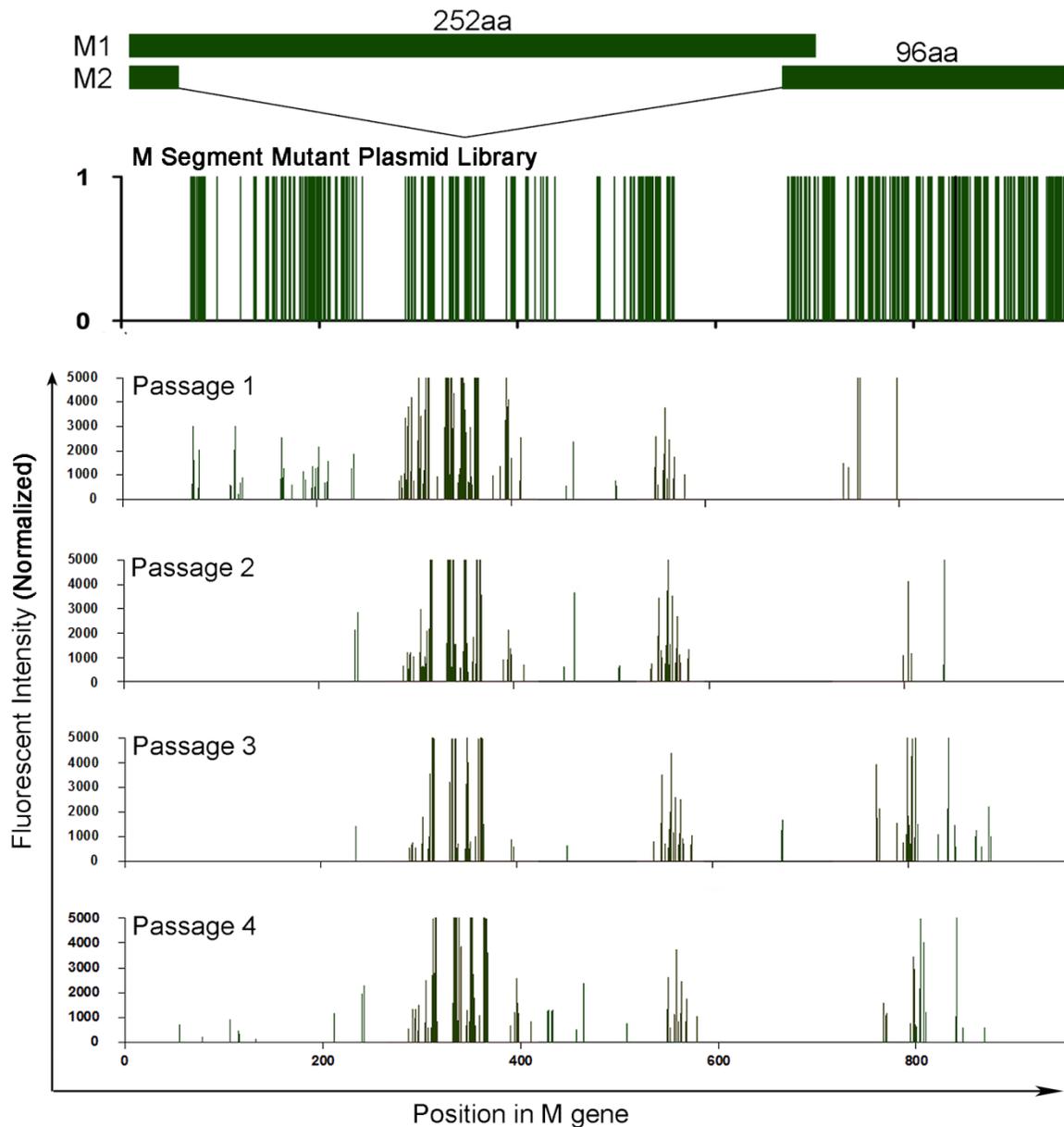
Supplementary Figure 2. Genotyping analysis of the M-gene mutant library. Schematic of influenza A WSN (H1N1) M1 and M2 proteins, with genotyping data displayed for the total mutant library (top); bars indicate insertion positions (nt) along the gene. The total mutant library was subsequently transfected into 293T cells and amplified multiple rounds in MDCK cells, labeled Passage 1-4. Peaks displaying fluorescence intensity indicate the amount of viral RNA. The normalization was performed for unselected pool M gene mutants as a negative control.

Fig. S3
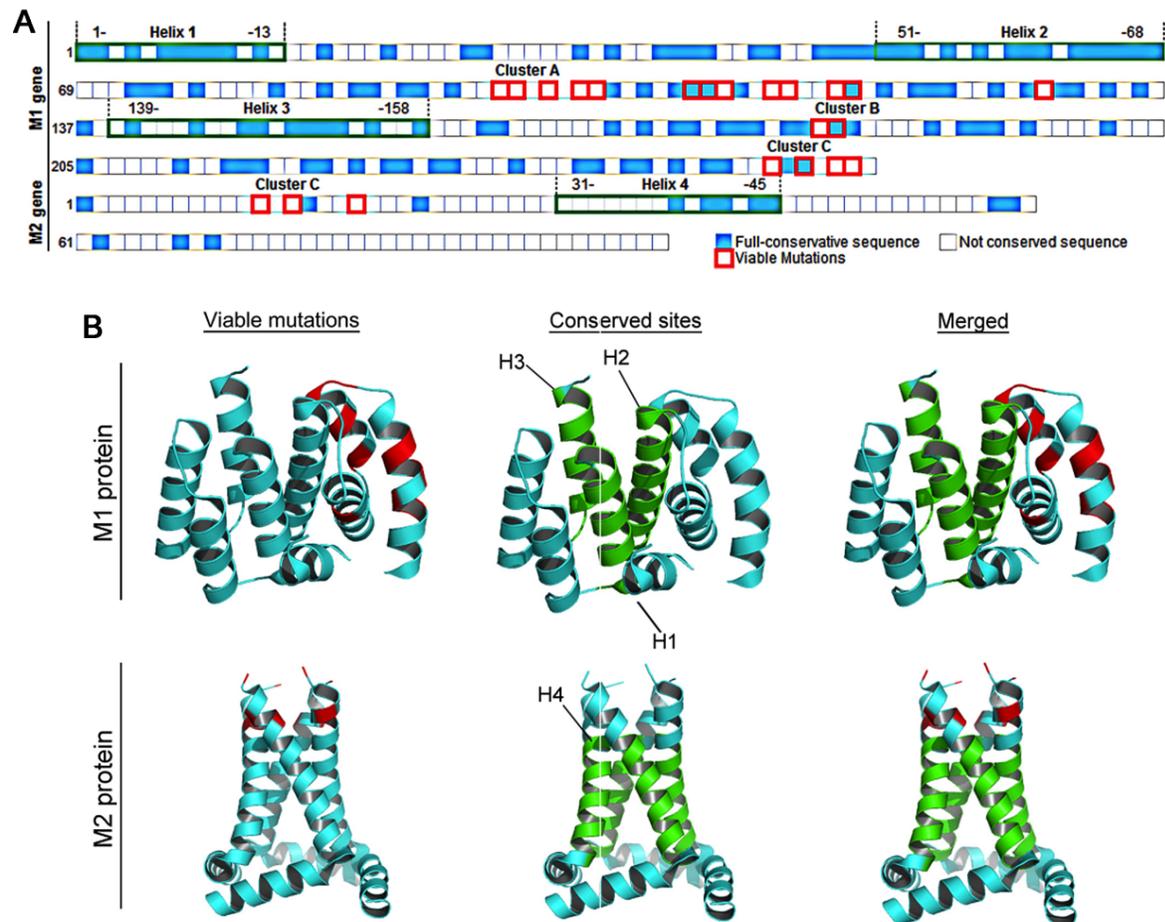
Supplementary Figure 3. Sequence and structure analysis of influenza matrix protein. **(A)** The protein sequences of influenza M1 and M2 proteins from approx. 300 strains published in the NCBI influenza database were compared by sequence alignment. Red boxes indicate clusters A, B, and C, identified through *in vitro* genotyping to be permissible to viral growth. Fully conserved amino acids in all known human virus H1N1 M1 and M2 genes are shown in blue, whereas those partially conserved are shown in white. Conserved helices are boxed in green (Helix 1,2,3 and 4). **(B)** Conserved and viable mutations in the M gene segment were mapped onto the crystal structure of the monomeric M1 gene (PDB accession ID, 2Z16) and the tetrameric M2 gene (PDB accession ID, 2L0J). Mutations permissive to viral growth, selected from the genotyping profile, are shown in red. Conserved sites (Helix1, 2, 3 and 4) from the M1 and M2 genes are shown in green.

Fig. S4.
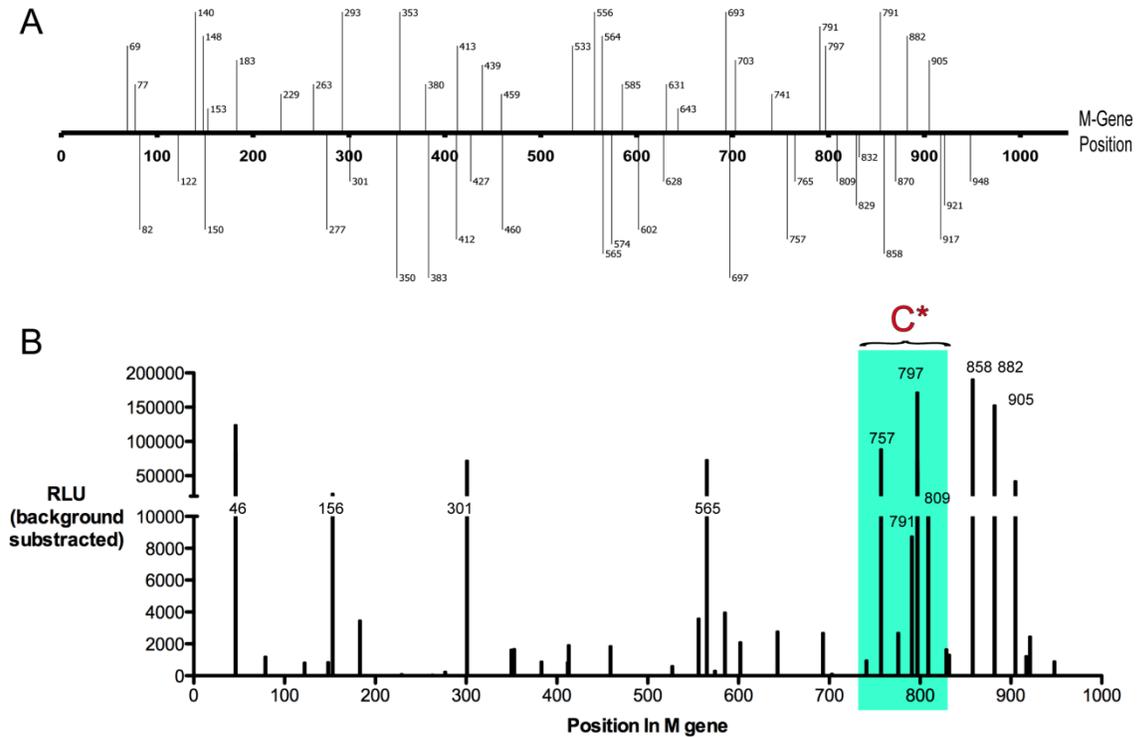
Supplementary Figure 4. Single clones selected from the M gene mutant library. (A) Schematic showing the positions of 67 individual mutant clones isolated across the M gene segment during initial mutagenesis. **(B)** The above mutant M plasmids were transfected with seven wt plasmids into 293T cells, and the supernatants were collected after 48hrs and re-infected onto a reporter cell line expressing Gaussia luciferase (gLuc) under a flu PA-promoter, which is an indirect measurement of infectious virus particles (see Methods). Highlighted in green are single clones located at the C* cluster, as shown in Figure 1A.

Fig. S5.

Influenza Infection in MDCK Cells
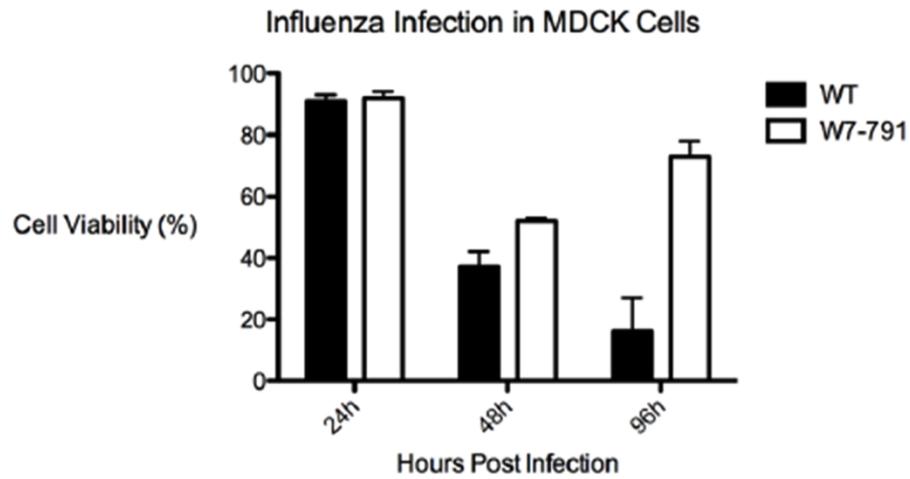
Supplementary Figure 5. W7-791 growth and *in vitro* toxicity. MDCK cells at 10^6 cells/well in a 6-well plate were infected with 0.1 MOI of wt WSN or W7-791. Viability of infected MDCK cells was determined at 24hrs, 48hrs and 96hrs post-infection using a LDH viability assay.

Fig. S6.

A

```
WT-M2        1 ATGAGTCTTCTAACCGAGGTCGAAACGCCTATCAGAAACGAATGGGGGTG        50
               |||||||||||||||||||||||||||||||||||||||||||||||||
W7-791       1 ATGAGTCTTCTAACCGAGGTCGAAACGCCTATCAGAAACGAATGGGGGTG        50

WT-M2       51 CAGATGCAACGATTCAAGTGATCCTC---------------TCGTCATTG        85
               ||||||||||||||||||||||||||               ||||||||||
W7-791      51 CAGATGCAACGATTCAAGTGATCCTCgtcattgcggccgcaTCGTCATTG       100

WT-M2       86 CAGCAAATATCATTGGAATCTTGCACTTGATATTGTGGATTCTTGATCGT       135
               ||||||||||||||||||||||||||||||||||||||||||||||||||
W7-791     101 CAGCAAATATCATTGGAATCTTGCACTTGATATTGTGGATTCTTGATCGT       150

WT-M2      136 CTTTTTTTCAAATGCATTTATCGTCGCCTTAAATACGGTTTGAAAAGAGG       185
               ||||||||||||||||||||||||||||||||||||||||||||||||||
W7-791     151 CTTTTTTTCAAATGCATTTATCGTCGCCTTAAATACGGTTTGAAAAGAGG       200

WT-M2      186 GCCTTCTACGGAAGGAGTGCCAGAGTCTATGAGGGAAGAATATCGAAAGG       235
               ||||||||||||||||||||||||||||||||||||||||||||||||||
W7-791     201 GCCTTCTACGGAAGGAGTGCCAGAGTCTATGAGGGAAGAATATCGAAAGG       250

WT-M2      236 AACAGCAGAATGCTGTGGATGTTGACGATGGTCATTTTGTCAACATAGAG       285
               ||||||||||||||||||||||||||||||||||||||||||||||||||
W7-791     251 AACAGCAGAATGCTGTGGATGTTGACGATGGTCATTTTGTCAACATAGAG       300

WT-M2      286 CTGGAGT     292
               |||||||
W7-791     301 CTGGAGT     307
```

B

```
WT-M2        1 MSLLTEVETPIRNEWGCRCNDSSDP-----LVIAANIIGILHLILWILDR        45
               |||||||||||||||||||||||||     :|||||||||||||||||||
W7-791       1 MSLLTEVETPIRNEWGCRCNDSSDPRHCGRIVIAANIIGILHLILWILDR        50

WT-M2       46 LFFKCIYRRLKYGLKRGPSTEGVPESMREEYRKEQQNAVDVDDGHFVNIE        95
               |||||||||||||||||||||||||||||||||||||||||||||||||
W7-791      51 LFFKCIYRRLKYGLKRGPSTEGVPESMREEYRKEQQNAVDVDDGHFVNIE       100

WT-M2       96 LE     97
               ||
W7-791     101 LE     102
```

Supplementary Figure 6. Nucleotide and protein sequence comparisons between wt WSN and W7-791. The nucleotide and protein sequences of the influenza M2 protein from wt WSN (WT-M2) was acquired from the NCBI influenza database. **(A)** Nucleotide sequence alignment between wt WSN and W7-791 **(B)** Protein sequence alignment between wt WSN and W7-791.

Fig. S7.
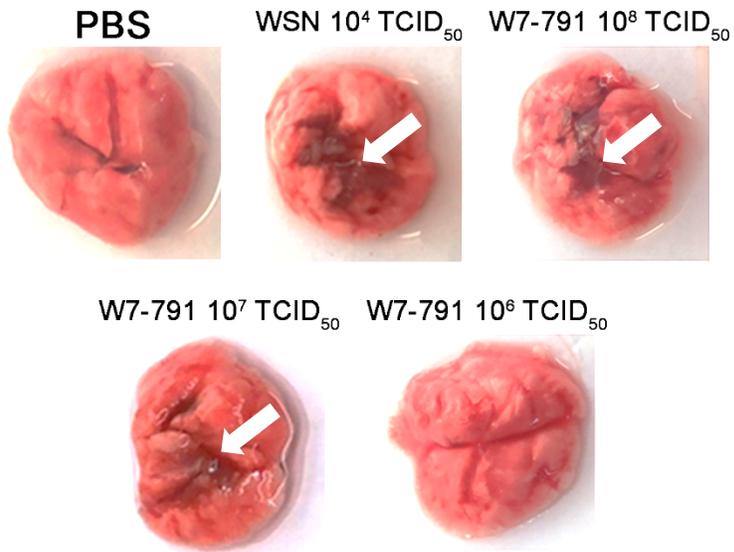
Supplementary Figure 7. Lung pathology of neonatal mice on day 4 post-inoculation.
Lung samples were isolated from 15-day-old mice after inoculation with PBS, WSN or the indicated titer of W7-791. Arrows indicate lung injury likely induced by the virus.
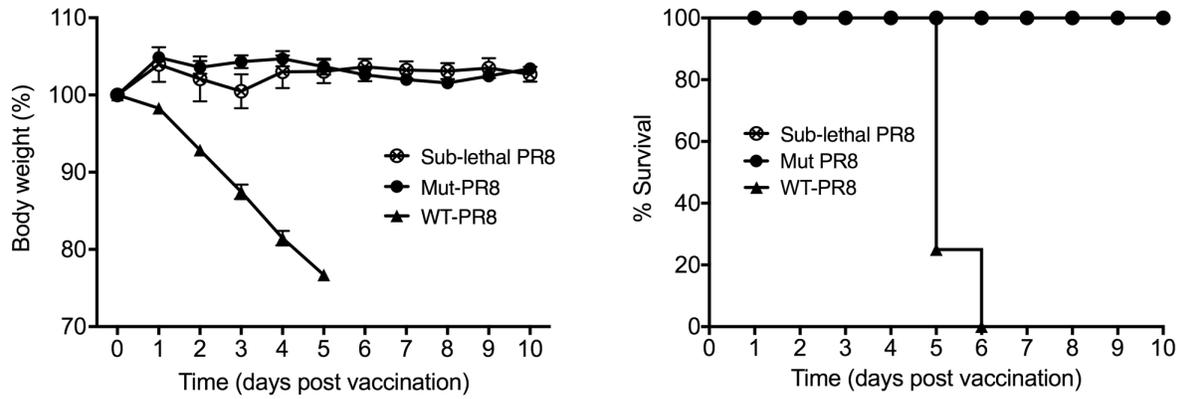
Supplementary Figure 8. The same insertion in PR8 attenuates the virus. B6 mice were immunized with 2000PFU of mut-PR8 or WT-PR8 or Sub-lethal dose of PR8 (250PFU) per mice, which is the equivalent of $4MLD_{50}$ of WT PR8 (n=5).

Fig. S9.
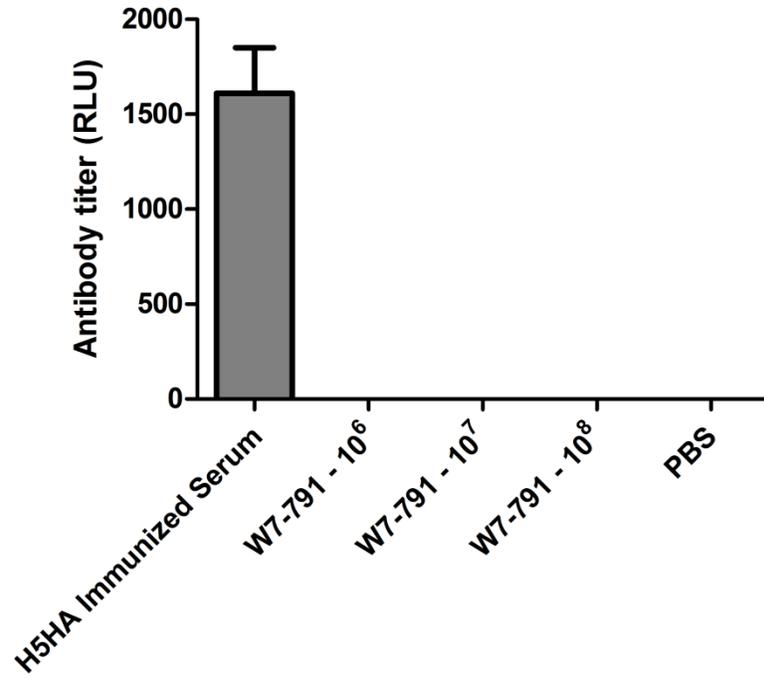
Supplementary Figure 9. W7-791 vaccinated ferrets do not produce detectable antibody against H5HA. The antibody titers of vaccinated and unvaccinated ferrets were measured using a CAM/H5 pseudotype virus neutralization assay. Serum from H5HA-DNA immunized ferrets was used as a positive control.

Fig. S10.

Ferret H3N1 Challenge

Ferret H1N1 Challenge
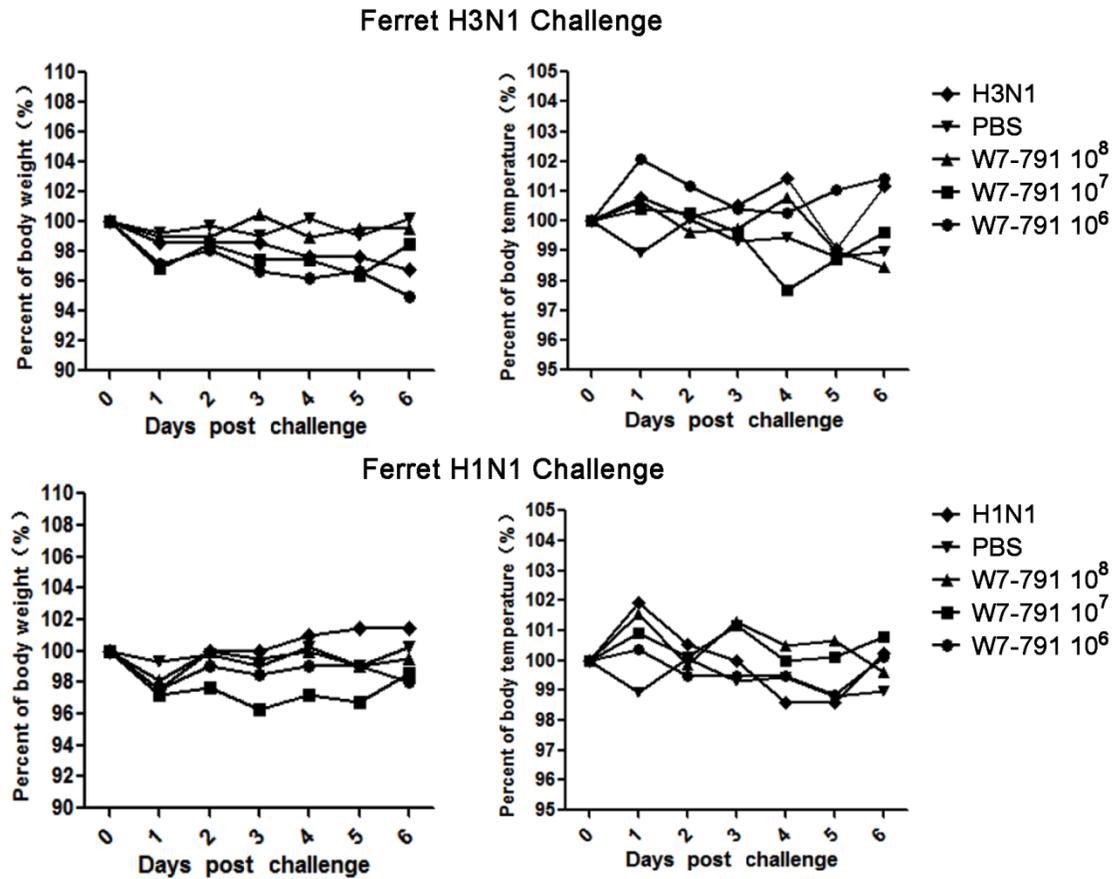
Supplementary Figure 10. No significant difference in the temperature or body weight of ferrets challenged with wt or W7-791 virus. The body weights and temperatures of ferrets challenged with WSN (H1N1) or HK68 (H3N1) are shown 6 days post H1N1 and H3N1 infection.

CHAPTER 4

CONCLUSIONS

The study presented in this thesis demonstrated rational designs to improve on several problems facing the field of virology today. First, we presented an alternative approach to determine the critical mutations during an outbreak of a lesser-known pathogen, ZIKV. Based on our analysis and predictions, we found four mutations in the prM gene which was linked to viral pathogenesis. Over the years, we have extended our zika research using single nucleotide polymorphisms, which provided us with more comprehensive tracking of viral mutations. During the course of research, we have also collaborated with hospitals and clinics from New York (United States) and Rio (Brazil) to obtain patient samples. This allows us to link patient medical history with phylogenetic and predictive modeling to provide clinically relevant and comprehensive Zika virus epidemiology since French Polynesia outbreak. We have further designed and tested a single primer pair as a ZIKV diagnostic tool to provide sufficient sequencing data to classify Zika virus. The results from this study have important implications for early determination of the nature of Zika virus. The working pipeline used in data analysis and prediction can also provide substantial information in the discovery of lesser-known pathogens.

Second, we have presented a rational approach to generate broadly neutralizing influenza vaccines. We have started with a viral pool of M-gene library, with each virus containing a single transposon insert. We then screened the entire viral pool *in vivo* and were analyzed at different timepoints using NGS. The data provided us with insights to the replication efficiency of influenza mutations at the M-gene, which eventually led to the discovery of W7-791, an influenza A vaccine candidate. Our challenge study revealed full protection against heterologous H1N1, H3N2 and highly pathogenic H5N1 in the mice model. At the time of this study, our vaccine, W7-791 has been the first candidate to provide cross-group influenza A virus protection in both mice and ferret with a single-dose

of administration. We have determined that the vaccine elicited both protective antibody and T cells using *in vivo* adoptive-transfer experiment. We further discovered that W7-791 provides improved protection when compared to seasonal quadrivalent influenza vaccine *in vivo*. These data suggested that W7-791 can elicit a robust immune response with a single administration. Our finding is signified as T cell specificity is now seen as the golden standard for the broad-spectrum anti-influenza protection. However, it is still unclear how W7-791 elicit T cell response efficiently. Future work to determine the immune mechanisms after administering W7-791 is substantial and may lead to the development of novel vaccines that takes advantage of the underlying immune mechanism in order for an individual to maximize the magnitude of the vaccine. Taken together, these findings provide a rational design for systematical screen for influenza virus vaccines, which could apply to any other virus with an existing reverse genetic system. Furthermore, globally interest to pursue a truly universal influenza vaccine or its mechanism will continue to be the main focus for influenza researchers.